ORACLE®

Oracle® Healthcare Translational Research

User's Guide

Release 3.2

E84083-02

September 2017

ORACLE®

Oracle Healthcare Translational Research User's Guide, Release 3.2

E84083-02

Contents

4 Working with the Cohort Viewer

5   Single Patient or Subject Viewer

6   Genomic Query

7 OBIEE Reports

8 User Management

Preface

This guide describes how to use Oracle Healthcare Translational Research (OHTR).

Documentation Accessibility

For information about Oracle's commitment to accessibility, visit the Oracle Accessibility Program website at http://www.oracle.com/pls/topic/lookup?ctx=acc&id=docacc.

Access to Oracle Support

Oracle customers that have purchased support have access to electronic support through My Oracle Support. For information, visit http://www.oracle.com/pls/topic/lookup?ctx=acc&id=info or visit http://www.oracle.com/pls/topic/lookup?ctx=acc&id=trs if you are hearing impaired.

Finding More Information

Oracle Help Center

The latest user documentation for Oracle Health Sciences products is available at http://docs.oracle.com/en/industries/health-sciences/.

My Oracle Support

The latest release notes, patches and white papers are on My Oracle Support (MOS) at https://support.oracle.com. For help with using MOS, see https://docs.oracle.com/cd/E74665_01/MOSHP/toc.htm.

Related Documents

Oracle Business Intelligence Enterprise Edition Documentation

The OBIEE documentation is available at http://docs.oracle.com/middleware/12212/biee/index.html.

Oracle Healthcare Foundation

The OHF documentation is available at https://docs.oracle.com/en/industries/health-sciences/oracle-healthcare-foundation/index.html

1

What You Can Do with Oracle Healthcare Translational Research

Use Oracle Healthcare Translational Research (OHTR) to:

- Select cohorts of patients or subjects using 10 clinical dimensions and 4 types of omics data type criteria.

- Save cohort queries and lists and share them with colleagues.

- View complete clinical and genomic information for a single patient or subject with the integrated Dalliance Genome browser.

- Use the built-in Visquick genome browser to view different types of genomic results in a circular plot.

- Search genomic data based on different genomic features like genes, genomic location, gene components, variants and so on.

- View information on patient or subject specimens linked to genes and variants.

- View and export specific clinical information for a selected cohort.

- View clinical activities on a visual timeline.

- View built-in cohort-driven reports on:

  - Demographic statistics

  - Clinical statistics

  - Genomic variants, copy number variation, gene expression and structural variants

  - Genes compared to a sample matrix

  - Variants compared to a sample matrix

- Export various genomic data of interest in formats like VCF, SEG, GCT and RES formats in either synchronous and asynchronous mode.

- Schedule jobs for genomic data export.

- Drill in from one screen to another to further drill down and generate various reports.

2

Getting Started

- Section 2.1, "Landing Page"

- Section 2.2, "Managing Queries"

- Section 2.3, "Managing Cohort Lists"

- Section 2.4, "Working with Gene Sets"

- Section 2.5, "Working with Jobs"

- Section 2.6, "Patient vs Subject Context"

> **Note:** If your encounter issues resulting in Internet Explorer 11 incompatibility, see the *Oracle® Healthcare Translational Research Installation Guide* for upgrade patch instructions to resolve them.

2.1 Landing Page

> **Tip:** For optimal viewing, ensure that your screen resolution is at least 1024 pixels.

Figure 2–1 My Workspace Screen
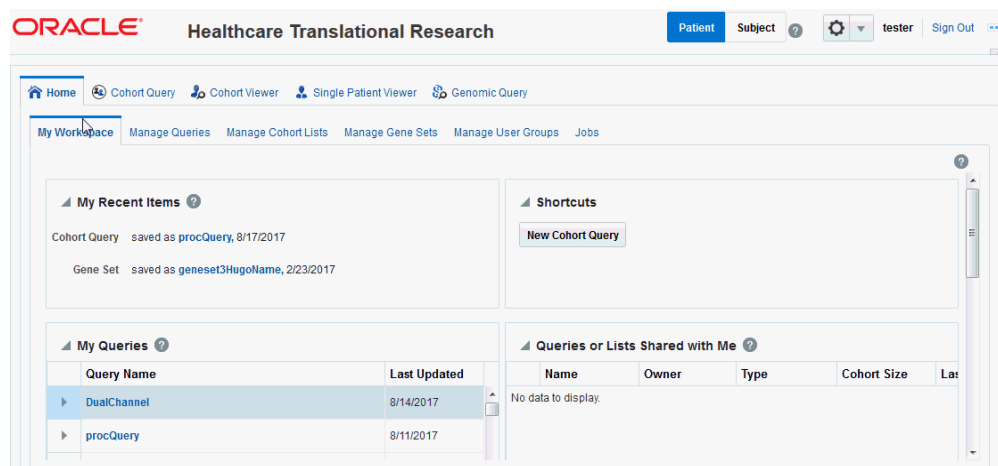
The **My Workspace** page is the landing page for OHTR. Depending on the license purchased, you may have access to most or all of the sections on this web page. With a

standalone OHTR license, you can only view the Cohort Queries and Cohort Lists details.

This page is a dashboard display of the recent items you have worked on, shortcuts and queries or lists shared with you by other users. Click the individual items on each page to view its details.

2.2 Managing Queries

The Manage Queries tab helps you:

- search through saved queries
- view a list of queries, either created by or shared with you
- modify queries
- set or modify restrictions to available queries

For details on modifying criteria for queries, inclusions and exclusions, see Chapter 3, "Working with Cohort Criteria and Queries".

2.2.1 Searching and Viewing Your Queries

You can only search for queries that you have access to. To search for a query:

1. Navigate to **Home** > **Manage Queries**.

2. In the **Search My Queries** section, enter the criteria to search for your query, for example, **Query Name**.

3. Click **Search**. The results are displayed in the **My Queries** section.

To clear the search and enter new criteria, click **Reset**.

4. To view a query, select the query and click **Load**.

2.2.2 Modifying Queries

Use the buttons at the top of the My Queries section to modify your queries.

2.2.2.1 Editing Query Name

To modify the query name:

1. Select the query in the **My Queries** section and click **Edit**.

   The Edit Query window is displayed.

2. Modify the query name and/or description.

3. Click **Submit**.

2.2.2.2 Deleting a Query

To delete a query:

1. Select the query in the **My Queries** section and click **Delete**.

   A confirmation window is displayed.

2. To delete the query, click **Submit**.

   If you do not want to delete the query, click **Cancel**.

2.2.2.3 Setting Restrictions on Queries

To set restrictions on who can access your queries:

1. Select the query in the **My Queries** section.

2. To ensure that no one else can access to the query, click **Private**.

3. To ensure that everyone else can access to the query, click **Shared With All**.

4. To share a private query, click **Shared**.

2.3 Managing Cohort Lists

The **Manage Cohort Lists** tab helps you:

- search through available saved cohort lists

- view a list of cohorts, either created by or shared with you

- modify cohort lists
- set or modify restrictions to available queries

2.3.1 Searching and Viewing Your Cohort Lists

You can only search for cohort lists that you have access to. To search for a cohort list:

1. Navigate to **Home** > **Manage Cohort Lists**.

2. In the **Search My Cohort Lists** section, enter the criteria to search for your saved cohort list, for example, **List Name**.

3. Click **Search**. The results are displayed in the **My Cohort Lists** section.

   To clear the search and enter new search criteria, click **Reset**.

4. To view a cohort list, select the cohort list in **My Cohort Lists** and click **Load**.

2.3.2 Modifying Cohort Lists

Use the buttons at the beginning of the **My Cohort Lists** section to modify your cohort lists.

2.3.2.1 Editing List Name

To modify the cohort list name:

1. Select the cohort list in the **My Cohort Lists** section and click **Edit**.

   The Edit Cohort List window is displayed.

2. Modify the list name and/or description.

3. Click **Submit**.

2.3.2.2 Deleting a List

To delete a cohort list:

1. Select the list in the **My Queries** section and click **Delete**.

   A confirmation window is displayed.

2. To delete the query, click **Submit**.

   If you do not want to delete the query, click **Cancel**.

2.3.2.3 Setting Restrictions on Cohort Lists

To set restrictions on who can access your cohort lists:

1. Select the list in the **My Cohort Lists** section.

2. To ensure that no one else can access the list, click **Private**.

3. To ensure that everyone else can access the list, click **Shared With All**.

4. To share a private query, click **Shared**.

2.4 Working with Gene Sets

Gene Set refers to groups or lists of genes. This set might comprise of a couple of genes or hundreds of genes. The concept of Gene Set lets you group genes into convenient

collections for reuse. You can group a few genes for quick search retrieval or for use in a cohort query.

- There are no restrictions on genes that can be included in a Gene Set. You can mix genes from multiple species or assembly versions.

- The same gene can be part of many different Gene Sets.

- Gene sets are private and cannot be shared among users.

- Gene set names are not case- sensitive.

- When genes are compared using the **Contains** or **Starts With** option, the limit is 512 characters.

2.4.1 Creating a New Gene Set

The create a new gene set:

1. Navigate to **Manage Gene Sets** > **Create New or Edit**.

2. Specify the genes you want to include in the gene set. This can be done in the following ways:

- **Type in Gene Names** - Search for genes using their Ensembl or HUGO names. You can enter multiple gene names separated by space, comma or semicolon.

- **Add from existing gene set** - Search for genes based on an existing Gene Set.

- **Upload from a file** - Select a text file from your desktop where the genes are delimited by comma, space or tab. Click **Choose File** to browse for the file. The file size limit is less than 5MB when genes are matched using Equals.

3. Specify **Species**.

4. Specify **Assembly Version**.

5. Click **Submit**. The matching genes appear in the **Available Genes** pane.

6. Use the arrow buttons to move available genes to the **Final Gene Set** panel. To remove genes from the Final Gene Set panel, use the **Remove** or **Remove All** buttons.

7. In the **Save Gene Set** section, enter a name for your gene set.

8. Enter a description, if desired.

9. Click **Submit**. The system saves the new Gene Set and a confirmation box appears.

2.4.2 Editing a Gene Set

You can add or remove genes from an existing gene set. The edit a gene set:

1. Navigate to **Manage Gene Sets** > **Create New or Edit**.

2. Search for the gene set you want to edit using the **Add from existing Gene Set** option.

3. Specify **Species**.

4. Specify **Assembly Version**.

5. Click **Submit**. The gene set appears in the **Available Genes** pane.

6. Specify the genes you want to include in the gene set. This can be done in the following ways:

■ **Type in Gene Names** - Search for genes using their Ensembl or HUGO names. You can enter multiple gene names separated by space, comma or semicolon.

■ **Add from existing gene set** - Search for genes based on an existing Gene Set.

- **Upload from a file** - Select a text file from your desktop where the genes are delimited by comma, space or tab. Click **Choose File** to browse for the file. The file size limit is less than 5MB when genes are matched using Equals.

7. Specify **Species**.

8. Specify **Assembly Version**.

9. Click **Submit**. The selected genes appears in the **Available Genes** pane.

10. Use the arrow buttons to move available genes to the **Final Gene Set** panel. To remove genes from the Final Gene Set panel, use the **Remove** or **Remove All** buttons.

11. Click **Submit**. The system saves the Gene Set and a confirmation box appears.

2.4.3 Managing a Gene Set

You can delete gene sets and view their metadata in this section. Gene Set metadata includes the name, description, owner of the gene set, privacy setting, the creation and updated dates and summary of how many genes are in a gene set.

To manage a gene set:

1. Navigate to **Manage Gene Sets** > **Manage**.

2. Enter the gene set you would like to modify.

3. Click **Submit**. Detailed information of the gene set is displayed.

4. To print the data, click **Print** on the top right.

5. To export the data, click **Export** at the bottom.

6. To delete a gene set, click **Delete Gene Set** at the bottom. A confirmation box appears to ensure you are performing the proper action.

   When you delete a gene set, its record is still available in the database but it can only be restored by a system administrator.

2.5 Working with Jobs

Currently, jobs are scheduled only for Genomic Data Export. To schedule a job, navigate to **Genomic Data Export** screen under the Cohort Viewer tab.

> **Note:**   Any job executed prior to applying OHTR 3.1 will have its status changed back to 'Scheduled'.

The following screen demonstrates how to create and schedule the job.

To create and schedule a job, perform the following steps:

1. Select the patient IDs source from one of the following options:

   - active query
   - library query
   - list
   - ad-hoc
   - Omics query

2. Select the DNA reference version.

3. Select the location for the gene from the following options:

   - Ad-hoc Gene List
   - Pathway
   - Gene Set

4. Select one of the following file formats to export:

   - Mutation-VCF
   - Copy Number Variation-SEG
   - Microarray Expression-RES
   - Microarray Expression Dual Channel-GCT

5. Select the **Schedule** export option. This prompts for the job name and description.

6. Click **Submit**. The job ID is created.

To view the job information and its progress, navigate to **My Workspace,** and click the **Jobs** tab. All the jobs created by the OHTR UI are displayed. Only the authorized user can see the job lists.

The following table describes the columns and buttons in the My Job page.

Table 2–1   Columns and Buttons in My Jobs Page

Column Name	Description
Job name	Job name specified when scheduling the job.
Job ID	Job ID generated while scheduling the job.
Job type	Only File Export job type is supported.
Scheduled on	Date and time on which the job is scheduled. The format is yyyy-mm-dd hh:mm:ss.
Repeat	You can schedule the job to run only once or recurrent.
Status	You can set the status to succeeded, failed, or canceled.

Table 2–1 (Cont.) Columns and Buttons in My Jobs Page

Column Name	Description
No. of times job executed	Count of how many times the job is run.
First execution started	Date and time when the job is scheduled to run. The format is yyyy-mm-dd hh:mm:ss.
Last execution successfully completed	Date and time when the job is last executed. The format is yyyy-mm-dd hh:mm:ss.
Search	You can search job by name or job ID.
Reset	You can reset your search criteria.
Delete	You can delete a job. A confirmation box appears to make sure you are performing the proper action. This is a soft delete, which means the job is still in the database but only the delete flag set to **Y**.
Abort	You can abort a job before it runs. You can also abort a queued or scheduled job but cannot abort a succeeded or canceled job.

2.5.1 Job Details

This page provides complete information about a single job.

- To view details for a particular job, from the Jobs page, click the link in the Job Names column.
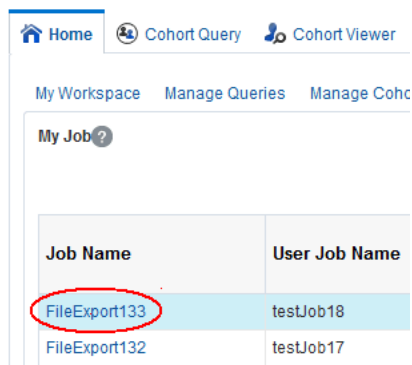
- To return to the main Jobs page, click the **Jobs** breadcrumb link.
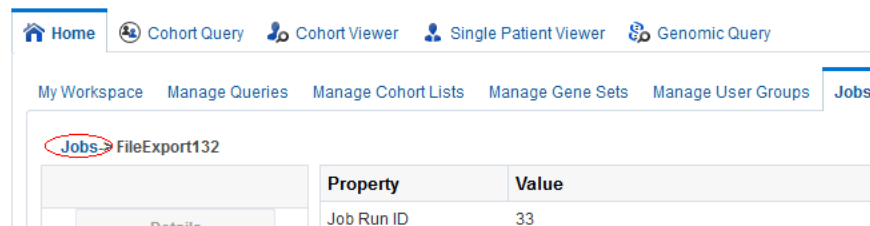
The following table describes job details.

Table 2–2 Job Details

Column Name	Description
Job run ID	Job instance ID generated when job is scheduled.
Job ID	Job ID generated while scheduling the job.
Job name	Name of the job.
Job type	Only File Export job type is supported.

Table 2–2   (Cont.) Job Details

Column Name	Description
Description	Description of the job.
Status	Status of the job (succeeded, failed, or canceled).
Owner	User who created the job.
Start date	Date and time the job is scheduled to run. The format is yyyy-mm-dd hh:mm:ss.
End date	Date and time when the job is last executed. The format is yyyy-mm-dd hh:mm:ss.

2.5.2 Job Inputs

This section provides the details of parameters provided when scheduling a job.

To view the details of the job created, click the **Jobs** tab and click **Inputs**.

The **Load input parameter values into Genomic Data Export page** link directs you to the genomic data screen with these values loaded. This occurs only when the job context matches the applications context. If not, the following note is displayed in place of link:

**Note: Selected Job cannot be loaded as its context is different from application context.

2.5.3 Job Outputs

This section shows the result files of a job. The following files are generated:

- Error log file - contains the log file of the application if any exception or error occurred while running the job.

- User log file - contains the log file, which shows the error occurred while generating the export files.

- Admin log file - is for DB administrator.

You can download error and user log files but cannot download the admin log file.

> **Note:**   When you try to export a file that does not have any Patients or Subjects linked to the specimen, an empty file is generated in Schedule mode.

2.6 Patient vs Subject Context

OHTR can be run in either Patient or Study Subject context. In patient context, all queries are directed at patient tables in Cohort Data Model schema, while in the study subject context, all queries are routed to the subject tables in CDM.
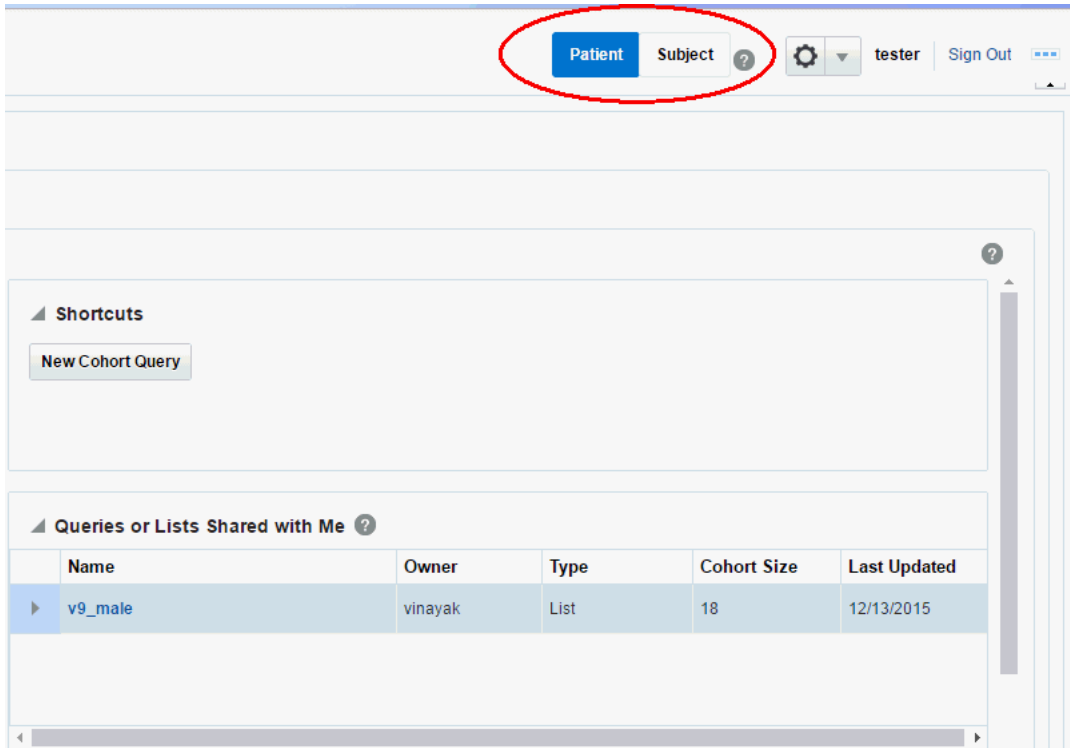
- By default, the context is set to Patient. You can change this to subject by clicking **Subject** on the upper right hand corner of the screen as shown above.

- The new context persists till the next time you change it.

- Switching context clears loaded data but not the criteria. Only the applicable criteria for the selected context is displayed. For example, Encounter criteria is not supported in the Subject context and will be disabled. When you switch back to the Patient context, it is enabled again.

- For a Subject to be considered deleted and not visible for selection, a given Subject's delete flag in CDM should be set to Y.

3

Working with Cohort Criteria and Queries

This chapter contains the following topics:

3.1 Categories and Topics

The query criteria are grouped into three broad data categories—Patient information, Clinical data, and Genomic data.

Each category is further divided into multiple topics where you can drill into specific parameters.

Table 3–1   Categories and Topics

Category	Topic
Patient Information	Demographics
Patient Information	Consent
Clinical Data	Diagnosis
Clinical Data	Clinical Encounter
Clinical Data	Procedure
Clinical Data	Medication
Clinical Data	Patient History
Clinical Data	Test or Observation
Clinical Data	Specimen
Clinical Data	Study
Clinical Data	Relative Time Events
Genomic Data	Sequence Variants
Genomic Data	Copy Number Variation

Table 3–1   (Cont.) Categories and Topics

Category	Topic
Genomic Data	Microarray Expression
Genomic Data	RNA-seq Expression
Genomic Data	Metadata Filters

3.1.1 Patient Information

3.1.1.1 Demographics

Figure 3–1   Patient Information Demographics options

Table 3–2   Demographics Details

Field Name	Definition	Sample Value or Values
Patient ID or Subject ID	Double-blinded unique identifier for the Patient. Also known as the Oracle ID	-
Gender	Patient's gender	Male, Female
Marital Status	Patient's marital status	Married, Single, Separated, Divorced
Age (in Years)	Patient's chronological age	-
Date of Birth	Patient's date of birth	-
Deceased Date	Patient's decease date	-
Ethnicity	Code to reflect Patient's ethnicity	-

Table 3–2 (Cont.) Demographics Details

Field Name	Definition	Sample Value or Values
Race	Code to reflect Patient's race	-
City	Code for Patient's City of residence	-
State	Code for Patient's State of residence	-
Zip Code	Code for Patient's Zip code	-
County	Code for Patient's County	-
Country	Code for Patient's Country	-

3.1.1.2 Consent

Figure 3–2 Consent

Table 3–3 Consent Screen Fields

Prompt Heading	Definition	Sample Value
Consent Type Code	Authorization for specified medical care.	Procedure Consent, Specimen Consent
Consent Status	The status of the consent form.	Active, Pending, Refused
Consent Start Date	The period start date of the patient's consent.	-
Consent End Date	The period end date of the patient's consent.	-

3.1.2 Clinical Data

3.1.2.1 Diagnosis

Figure 3–3   Diagnosis

Click the magnifying glass icon for **Diagnosis,** to search either by Diagnosis Code or Diagnosis Name.

Table 3–4   Diagnosis Screen Fields

Prompt Heading	Definition	Sample Value
Diagnosis	Code that classifies the patient's clinical condition.	Most commonly identified with ICD codes.
Diagnosis Status	Code that reflects the status of the Diagnosis	Active, New, Recurring
Onset Date	Date of the onset	-
Reported Date	Date when the diagnosis was recorded by a service provider	-
End Date	Date of resolution.	-
Age at First Onset	Age of the patient at first onset (years)	-
Anatomical Site	Anatomical site or sites related to the diagnosis	-

3.1.2.2 Clinical Encounter

Figure 3–4 Clinical Encounter

Click the magnifying glass icon, next to Encounter Type, to search either by Encounter Code or Encounter Name.

Table 3–5 Clinical Encounter Screen Fields

Prompt Heading	Definition	Sample Value
Encounter Type	Type of clinical encounter that individual has undergone, valid values may be inpatient, outpatient	Inpatient, Outpatient
Location	Facility where it took place, generally name of the hospital, clinic, doctors office and so on	Sequoia Hospital, Medical Associates, Mass General Hospital Ear and Nose Dept
Time	Date when encounter took place	-
Datasource	Name of the data system actual clinical data is coming from	EMR1, EMR2

3.1.2.3 Procedure

Figure 3–5   Procedure

Click the magnifying glass icon for **Procedure,** to search either by Procedure Code or Procedure Name.

Table 3–6   Procedure Screen Fields

Prompt Heading	Definition	Sample Value
Procedure	A discreet intervention performed by a clinician. Procedures are most commonly identified with CPT codes	-
Procedure Type	A sub categorization of procedures	Surgical, Radiology, Diagnostic
Procedure Start Date	The date when the procedure began	-
Procedure End Date	The date when the procedure concluded	-

3.1.2.4 Medication

Figure 3–6   Medication

Click the magnifying glass icon for **Medication** to search either by Medication Code or Medication Name. There are two search modes available. For details, see Medication Search Modes.

Table 3–7   Medication Screen Fields

Prompt Heading	Definition	Sample Value
Medication	A pharmaceutical substance intended to provide therapeutic benefit.	NDC codes, RxNorm codes
Medication Start Date	The start date for the administration of the medication.	-
Medication End Date	The end date for the administration of the medication.	-
Dosage	The medication's dosage.	-
Dosage Units	Code for the medication dosage's units.	-
Medication Outcome	Clinical outcome of taking a given medication as assessed by clinician or medical professional	Partial Recovery, Full Recovery, Adverse Reaction

3.1.2.4.1   Medication Search Modes

The Medication search menu supports the following two search modes:

Classic Search

Classic search lets you search for Medication codes based on medication name, medication code and code system. These parameters can be used with the following options:

- Starts with

- Ends with

- Equals

- Does not equal

- Contains (this is the default value)

- Does not contain

- Is blank

- Is not blank

To perform a search in classic mode:

1. Navigate to **Cohort Query** > **Clinical Data** > **Medication**.

2. Click the magnifying glass icon next to the **Medication** field.

3. Select **Classic Search**.

4. Specify the search criteria either by using the above parameters individually or in combination, along with the available operators.

Figure 3–7   Classic Search

| Search Mode | ◉ Classic Search |
| | ◯ Hierarchy Drill-in Search |

Medication Name	Contains ▾	
Medication Code	Contains ▾	
Code System	Equals ▾	▾

Search   Reset

The search is case insensitive

Code	Name	Code System
No data to display.		

Code	Name	Code System
No data to display		

»

Remove   Remove All

Cancel   Submit

5. Click **Search**.

6. The Code, Code Name and Code System of the matching codes are displayed on the left. Expand the list items to view details about code hierarchy and the Code Description.

If you have not specified any search criteria, then all codes at all levels of hierarchy from all the hierarchies are listed.

7. Select any number of rows on the left and copy them to the right hand side selection list using the **>** button.

8. Click **Submit** at the bottom of the screen. All the selected codes are used for defining query criteria in the Cohort Query.

Hierarchy Drill-in Search

Hierarchy Drill-in search lets you search for Medication codes based on medication name, medication code and code system. These parameters can be used with the options:

- Starts with

- Equals

- Contains (this is the default value)

1. Navigate to **Cohort Query** > **Clinical Data** > **Medication**.

2. Click the magnifying glass icon next to the **Medication** field.

3. Select **Hierarchy Drill-in Search**.

Figure 3–8   Hierarchy Drill-in Search

4. Specify the search criteria either by using the above parameters individually or in combination, along with the available operators.

   The drop down list for **Code System** lets you select multiple Code Systems.

5. Click **Search**.

6. The matching codes are displayed on the left. The entire hierarchy in which the matching code is present, is listed as a hyperlink. Click the hyperlink to view an indented list where each child level code is displayed at one indent level more than its parent code.

   If you have not specified any search criteria, then all the codes at all the levels of hierarchy, from all the hierarchies are listed.

Figure 3–9   Matching Codes from Hierarchy Drill-in Search

7. To view the count of patients who were given the medication represented by the Code, select **Code(s)** and click **Count Patients**. The count is cumulative of the selected Code and all its child Codes.

   For example, if Second Level Child Code is selected, then the count corresponds to the Patients who were given the Medication represented by Second Level Child Code or Medications represented by any of the children of the Second Level Child Code.

8. Select any number of rows on the left and copy them to the right hand side selection list using the > button.

9. Click **Submit**. All the selected codes are used for defining query criteria in the Cohort Query.

3.1.2.5 Patient History

Figure 3–10   Patient History

Click the magnifying glass icon for **Patient History** to search either by Medication Code or Medication Name.

> **Note:**   In the **Patient History** dialog, the History Type by default uses the Equals operator and lists the first 100 history types only. If there are more than 100 history type values in your data, you can search using other operators like "Starts with," "Contains," and so on.

Table 3–8   Patient History Screen Fields

Prompt Heading	Definition	Sample Value
Patient History	A coded representation of the Patient History	Smoking, Obesity
Patient History Start Date	Date when this history was known to be active	-
Patient History End Date	Date when this history is known to no longer be active	-
Amount	Numerical value	For example, 1
Amount Units	Unit of Measure	For example, mmHg
Frequency	Numerical Value	For example, 12
Frequency Units	Unit of Measure	For example, TPDAY (times per day)

Table 3–8   (Cont.) Patient History Screen Fields

Prompt Heading	Definition	Sample Value
Text Value or Code	Value of history which is represented by text or coded.	Yes, No, Frequently, Rarely
History Applicable To	For Familial History, this will contain type of blood relative, which may have a history value	Father, Mother, Paternal grandmother, Paternal Grandfather

3.1.2.6 Test or Observation

Figure 3–11   Test or Observation
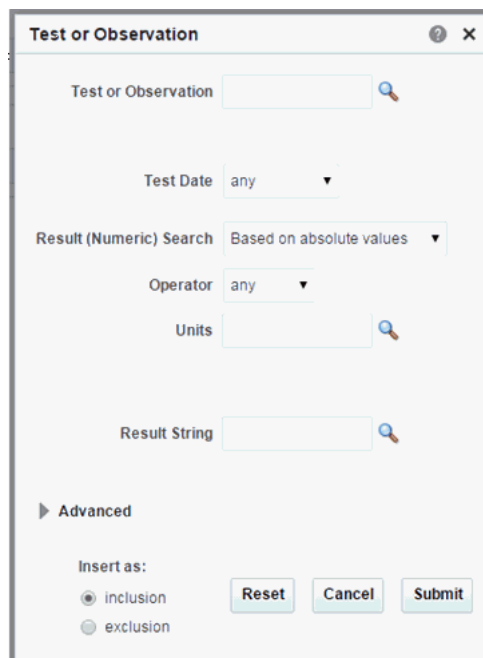
Click the magnifying glass for **Test or Observation** to search either by Test or Observation Code, Test or Observation Name. The units will be filtered based on the selected Test or Observation.

> **Note:**   In the **Search Test or Observation** dialog, the Test or Observation Type by default uses the Equals operator and lists the first 100 types only. If there are more than 100 Test or Observation type values in your data, you can search using other operators like "Starts with," "Contains," and so on.

Table 3–9   Test or Observation Screen Fields

Prompt Heading	Definition	Sample Value
Test or Observation	Any kind of medical intervention performed to aid in the diagnosis or detection of disease	Blood Pressure
Test Date	The date the test was performed	May-12-2012

Table 3–9   (Cont.) Test or Observation Screen Fields

Prompt Heading	Definition	Sample Value
Result (numeric) Search	The Test result	140, 90
Operator	This search mode enables searching numeric results with UoMs based on reference range or absolute values.	
Units	The Test units of measure	mmHg
Result String (text)	Textual Test result, also Notes or remarks for the Test	Positive/Negative

3.1.2.7 Specimen

Figure 3–12   Specimen

> **Note:**   The three fields of and **Specimen Type**, **Specimen Number**, **Specimen Vendor** are used in combination to search for particular data. Click the magnifying glass icon next to each field, to specify the appropriate criteria.

Table 3–10   Specimen Screen Fields

Prompt Heading	Definition	Sample Value
Specimen Type Name	A coded description for the type of Specimen.	Blood, Urine, Sputum
Specimen Number	A unique identifier for a specimen.	-

Table 3–10   (Cont.) Specimen Screen Fields

Prompt Heading	Definition	Sample Value
Specimen Vendor Number	A unique identifier for a source of the specimen, namely lab the specimen was analyzed.	Harvard-CancerInstitute01, MIT-Whitehead-Lodish5
Anatomical Site	The target site of the intervention.	-
Specimen Collection Date	Date the specimen was collected.	-
Specimen Amount	The amount of the specimen collected.	-
Units	The units of measure for the specimen collected.	Mg/dL, millimoles/liter

3.1.2.8 Study

Figure 3–13   Study Screen
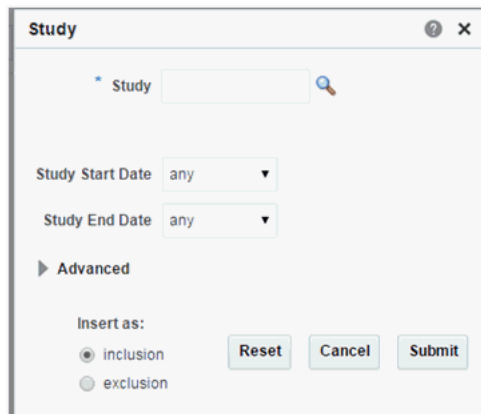
Click the magnifying glass icon for **Study,** to search either by Study Code or Study Name.

Table 3–11   Study Screen Fields

Prompt Heading	Definition	Sample Value
Study (Name or Identifier)	A reference to a particular study.	NOT-A-STUDY (any genomic data that is not explicitly tied to study would be categorized there), Glioblastoma Study
Study Start Date	The start date for the study.	-
Study End Date	The end date for the study.	-

3.1.2.9 Relative Time Events

Frequently, the events that drive cohort or patient set searches are relevant to finding patients that may have certain event dependencies such as being diagnosed with a disease after taking a medication.

You can search for such patients based on relative time event dependencies. You can set this criteria and search for patients who have a history of one clinical (or genomic change) event relative to another.

In other words, you can specify a search for patients that took a certain medication or a set of medications within 30 days before being diagnosed with a specific disease. You can also search for patients who had a procedure performed on them two months after starting a medication.

Figure 3–14   Relative Time Events

To select relative time events:

1. Specify the event data to start your event.

2. Specify the relative time condition. This is how the events in Step 1 and Step 3 relate to each other.

3. Specify the related event to associate with the event in Step 1.

The search criteria is interpreted - Search for patients that experienced Event X at any time while, certain amount of time before or after Event Y. The idea is that X is the original event, Y is the associated event, and the time frame describes the time relationship between the two events, for the given patient.

The events listed for selection in Step 1 or Step 3 are present in Clinical or Genomics Data categories.

However, it is important to explain what the Relative time conditions are in Step 2 and how their meaning translates into SQL:

■ *at any time while* - the meaning of this relative time condition is to find any events that overlap with each other in terms of their occurrence. For example, if we have two distinct events with start and end dates, as long as at some point in time both events were true, this condition is satisfied.

Example 1:

Event in Step 1 - Started on May 1, 2010 - ended on June 29, 2011

Event in Step 3 - Started on June 10, 2011 - till present

The two events have overlap between June 10-29, 2011 thus the condition is satisfied.

Example 2:

Event in Step 1 - Started on Dec 1, 2010 - ended on Dec 12, 2011

Event in Step 3 - was performed on August 3, 2011

The two events have overlap on August 3, 2011 thus the condition is satisfied.

Example 3:

Event in Step 1 - Started on May 1, 2010 - ended on June 29, 2011

Event in Step 3 - Started on June 10, 2009 - till present

The two events have overlap between May 1, 2010 and June 29, 2011 thus the condition is satisfied.

The dates used for time comparing each event type are as follows:

– Diagnosis: Reported Date - End Date (if null, use present date)

– Procedure: Start Date - End Date (if null, use present date)

– Medication: Start Date - End Date (if null, use present date)

– Test or Observation: Test Date (single day event only)

– Gene Variant: Specimen Collection Date for end date use present date

■ *(more than / less than / exactly) [ ] (days / weeks / months / years) (before / after)* - the meaning of this relative time condition is to find any events that have start or occurrence dates that relate to each other with the time period specified by the condition.

Examples:

Start date of Event in Step 1 is more than 5 weeks before Start date of Event in Step 3

Occurrence, for example, Test Date of Test or Observation in Step 1 is less than 1 month before Start Date of Medication in Step 3

Gene Variant Specimen Collection Date in Step 1 is exactly 1 day after Procedure Start Date in Step 3

The dates used for time comparing each event type are as follows:

– Diagnosis: Reported Date

– Procedure: Start Date

– Medication: Start Date

– Test or Observation: Test Date

– Gene Variant: Specimen Collection Date

3.1.3 Genomic Data

3.1.3.1 Microarray Expression

Microarray expression is one of the growing set of genomic criteria that can be used to stratify patients. The genomic data driving this filter comes from the Oracle Healthcare Omics (OHO, formerly known as ODB) model, specifically the W_EHA_RSLT_GENE_ EXPR table and related tables. Gene expression data is associated with patients through the specimen used for genomic study analyses collected during study participation or other genomic testing.

The upper section of the criteria gives you the choice to specify Specimen Type and Anatomical site to further filter the results.

In the popup for Gene Expression, you are given two options (Array data types to choose from). For each option, you must specify the criteria when selecting patients based on their gene expression data.

■ **One-channel** – Single channel data can be filtered based on the following data fields:

– On Intensity; as a cutoff based on aggregates or on values, and across an experiment or multiple ones.

– P-value; as a cutoff on values.

– Call; on any of the three call types.

■ **Two-channel** – Here data can be filtered on the following data field:

– Log2Ratio – as a cutoff on values

In **Expression for Genes From**, select at least one unique gene from any of the three specified sources:

1. Ad-hoc list

2. Pathway

3. Gene Set

Figure 3–15   Microarray Expression

Click the magnifying glass icon for **Specimen Type**, to search either by Specimen Name or Specimen Code. This also applies to the other search criteria. Criteria identified with * are required.

Table 3–12   Microarray Expression Screen Fields

Prompt Heading	Definition	Sample Value
Specimen Type	The specimen type description that identifies the specimen used for genomic tests.	Normal sample, tumor sample.
Anatomical Site	Anatomical site name or code corresponding to the specimen collected.	Left lung, kidney, small intestine

Table 3–12   (Cont.) Microarray Expression Screen Fields

Prompt Heading	Definition	Sample Value
Intensity	This and the following two fields are specific to Single Channel data. This field is a Gene expression value range specification. Data driving this selection is in the W_EHA_RSLT_ GENE_EXP table in OHO and is generally expected to be normalized now to take full advantage of this criteria selection. Additionally, the data values should all be positive. The criteria conditions allow you to narrow down patients based on up-regulated gene expression (using *greater than 1.0 times mean* condition either across columns (*same hybridization*) or rows (*single result file*) of the data in the particular experiment.	Down regulated gene in hybridization example - Intensity is less than 2.0 times mean in the gene expression within a particular hybridization.
P-value	Significance value associated with the specific experiment, optional.	P-value < 0.00001
Call	Call made on the particular value, if present. All are taken unless specified. Optional.	P - Present, A - Absent, M - Marginal
Log2Ratio	Specific to dual channel differential gene expression relating the difference between, for example, control sample and the tested sample intensity. Data for this selection in found in the **W_EHA_ RSLT_2CHANNEL_GXP** table.	Float type integer representing a log base 2integer value of a fraction.
Expression for genes from*	Selection of genes that are to be used for patient stratification based on the expression. At least one of the below criteria must be specified.	N/A
Ad-hoc List: Gene	List of one or more genes.	-
Pathway	Reference to a pathway stored on the reference side of the Oracle Healthcare Omics (OHO, formerly known as ODB) model. This in turn corresponds to a list of genes that are to be used to compare their Intensity values.	-
Gene Set	User-defined collection of genes that you can reference across any UI instead of having to build ad-hoc lists of genes each time.	MyGeneSet1, GlioblastomaSmithLabGenes
Assembly Version	Represents GRCh assembly version. Default selection is last loaded version.	GRCh38 or GRCh37
DNA Reference Version	Represents the Ensembl reference version for getting gene annotations. This is related to 'Assembly Version' and by default it shows the preferred DNA Reference Version which is set in OHO.	Ensembl release 70

For Metadata Filter details for Microarray Expression, see Metadata Filters on page 3-30.

3.1.3.2 Sequence Variants

The Variants Criteria Selection interface falls under the category of Genomic criteria and enable the user to query for patients based on results present in the **W_EHA_ RSLT_SEQUENCING** and related tables in the Oracle Healthcare Omics (OHO, formerly known as ODB) model.

In the Sequence Variant screen, you have the option to select Specimen Type and Anatomical site for associating the selected variant criteria in this screen. Currently, the following 5 main options are available for searching:

- having Variants in selected Genes: this enables you to search variants in specific genes.

- having selected Genomic Variants: this enables you to search for specific known variants by their Cosmic or dbSNP identifiers.

- having Variants within specified Genomic Region: this enables you to search variants in specified genomic location like chr1:19094593-29302393 or whole of chr1.

- having Zygosity: this enables you to search variants with specific zygosity.

- having Genotypes: this enables you to search for specific genotypes like AT, or AA, or wildtype (same as reference), and so on.

Additionally there are other parameter, attributes and metrics associated with variants that you can use for further filtering on specific variants. These options are:

- Specifying variants by their attributes.

- Specifying variants by their non-synonymous substitution scores

- Specifying quality metrics filters depending on the sequencing file type like VCF, MAF and CGI masterVar.

The filtering options have been categorized into 2 steps. Step 1 as shown in Figure 3–16 has all mandatory filter options for creating a query. Step 2 shown in Figure 3–17 has the optional filters.

Figure 3–16 Mandatory Filters

Figure 3–17 Optional Filters

Table 3–13 Variants Screen Fields

Field Name	Definition	Sample Value
Specimen Type, Anatomical Site	For more information on Microarray Expression, refer to description under Microarray Expression section on page 3-17.	As in Microarray Expression section.
Having Variants in selected Genes	Variant mode selection - being able to specify variants in specified genes	N/A
Having selected Genomic Variants	Variant mode selection - being able to specify known variants by their identifier. At least one mode of variant specification must be given.	N/A
having Variants within Specified Genomic Region	Variant mode selection - being able to specify variants in a specific genomic location like chr1:1234324-3434333, chr7:1000, chr3 or chr2:1000+200	N/A
Genomic Variant ID	Specify known gene variant reference identifiers such as dbSNP or Cosmic. Allows for selecting multiple values.	rs56289060, 905944
	Assembly Version helps specify variants in a selected assembly version. Enables multiple selection, except for 'having Genotype' option where only single selection is allowed.	
DNA Reference Assembly Version	Helps specify variants in a selected reference assembly version. Allows multiple selection, except for *having Genotype* option where only single selection is allowed.	GRCh37

Table 3–13   (Cont.) Variants Screen Fields

Field Name	Definition	Sample Value
DNA Reference Version	Represents the Ensembl reference version for getting gene annotations. This is related to 'Assembly Version' and by default displays the preferred DNA Reference Version, which is set in Oracle Healthcare Omics (OHO, formerly known as ODB).	Ensembl release 70
At Genomic Position	Helps specify a genomic region in which variants has to be searched.	chr7, chr7:1000, chr3:1000-2000, chr2:1000+200
Genotype	List available genotype values for the selected genomic position or genomic variant based on the position. Also displays wildtype base (same as reference base). User can select a combination of two genotypes to search or just one genotype and search for patients with these selected genotypes.	N/A
Additional Variant Information (Optional) are used in the context of either of the radio buttons selection above	Additional criteria used to filter down variants based on variant type and variant impact features.	N/A
Variant Type and corresponding Variant Impact	You can select any of the variant types supported along with additional variant metadata such as each variant type's impact on the resulting protein. **Note:** Not all Variant Impact annotations apply to each Variant Type.	Variant Type: Substitution, Insertion, Deletion, Indel, Complex Variant Impact: Synonymous, Missense, Nonsense, Unknown,
Variant Status	Specifies which variant types to consider - whether the variant should be known or novel. Default considers all variants.	Known, Novel
Strand	Gene transcription direction attribute. By default, all directions of transcription are included.	+ means forward, - means reverse
Ad-hoc List: Gene; Pathway; Gene Set	For more information on Microarray Expression, refer to description under Microarray Expression section on page 3-17.	As in Microarray Expression section.
At Genomic Position	Specify genomic location for the variants to occur in, the format should be chr#:from-to or chr# if entire chromosome to be used.	chrX:13000-120000, chr7

Table 3–13   (Cont.) Variants Screen Fields

Field Name	Definition	Sample Value
Non-synonymous Substitution Scores	Data for this section is loaded into the reference side of the OHO model from Ensembl based on either PolyPhen algorithm or SIFT algorithm. The prediction value can be specified numerically or alternatively can be specified using PolyPhen or SIFT specific annotation. **Note**: SIFT or PolyPhen predictions are only available for Known variants	With PolyPhen, prediction between 0 and 1 or labeled as benign, Probably Damaging, possibly damaging and unknown With SIFT, prediction between 0 and 1 or labeled as Deleterious and tolerated
Variant Parameters Depending on Sequencing File Type	User can select to specify more detailed filtering criteria based on data coming from 3 different sequencing file formats such as VCF, MAF, CGI masterVar. As each input file formats uses different metadata to describe stored entities, depending on the sequencing input format selection, the user can elect to specify: VCF: Variant Call Format, Format.GQ range - user can specify upper or lower numeric values for this parameter MAF: Mutation Annotation Format, Score - user can specify upper or lower numeric values for this parameter Somatic Status Somatic Score Allele Read Count Reference Read Count Total Read Count RMS Base Quality RMS Mapping Quality AD/DP Ratio CGI masterVar: Complete Genomics masterVariation format. The available fields to search by are ■ Allele Zygosity ■ Score VAF ■ Score EAF ■ Allele read count ■ Reference read count	See appropriate file formats documentation for appropriate value ranges (VCF 4.2 format, MAF 2.0-2.2 format, Complete Genomics masterVar format). For example, Allele Zygosity for CG masterVar includes het-alt, hom, half, het-ref options.

For Metadata Filter details for Microarray Expression, see

3.1.3.2.1   At Genomic Position

You can opt to specify genomic data selection using genomic co-ordinates. You specify the chromosome region in a standard format for the Variation and CNV data. You can

specify a complete chromosome or part of a chromosome as criteria. Currently, only one chromosome region at a time is implemented for search.

The following chromosome region formats are supported.

- CHR15:10000-200000: Considers region between 10000 to 200000 in chromosome 15.

- CHR15:1,200,000+5000 - Considers 5000 bases upstream from 1,200,000 position in chromosome 15.

- CHR15 - Considers whole of the chromosome 15.

- CHR15:1000 - Considers 1000th nucleotide position of chromosome 15.

3.1.3.3 Copy Number Variation

The Copy Number Variation criteria selection interface falls under the category of Genomic criteria where you can query for patients based on results present in the **W_EHA_RSLT_COPY_NBR_VAR** table and related tables in the Oracle Healthcare Omics (OHO, formerly known as ODB) model. Currently, this table contains data from two formats—from Complete Genomics and SEG and VCF files with CNV data. Both of these have log2 ratio stored in OHO.

For CNV, you can optionally select specimen type and anatomical site.

Next select CNV Result Type, which represents data from numeric (array based) and categorized sequencing based platform.

You can then filter results based on the list of Copy Number Variation Attributes. For example, for numeric based CNV Result Type, SNP log2 Ratio and for categorized based, gain, loss, equal to indicate Amplification, Deletion or no change in the copy number of a given gene or gene region.

Finally, you should specify the location of Copy Number Variation which is the gene or genomic position of interest.

Figure 3–18   Copy Number Variation - Numeric Based

Figure 3–19　Copy Number Variation - Categorized Based

Table 3–14　Copy Number Variation

Prompt Heading	Definition	Sample Value or Values
Study*, Specimen Type, Anatomical Site	For more information on Microarray Expression, refer to description under Microarray Expression section on page 3-17.	As in Microarray Expression section.
CNV Result Type	Search CNV results either belonging to array based platform like Genome_Wide_SNP_6 array or sequencing based CNV data from complete Genomics.	Array based or sequencing based
SNP Log2 Ratio (Segment Mean)	Values for segment mean in the form of a range. You can also specify a single value in Log2 Ratio and search for results with segment mean greater than the specified value.	Numeric value. It can accept negative values.
CNV Type	Copy Number Variation attribute indicating whether it is an amplification - gain, deletion - loss, or no change - equal.	Gain, Loss, Equal
With confidence > (CNV Type score)	Copy number variation confidence score associated with CNV Type. Score is populated from the source file, and depending on the scoring method, the range can vary.	Numeric value, range can vary depending on source

Table 3–14   (Cont.) Copy Number Variation

Prompt Heading	Definition	Sample Value or Values
Called Ploidy	Values for ploidy can be given more specifically as a range, either upper and/or lower bound can be specified. For example, for duplication, called ploidy can be specified as 2.	Range of Ploidy to be selecting based on, e.g. for duplication, it can be specified as between 1.5 and 2.5
With confidence > (Ploidy score)	Confidence score associated with Called Ploidy. Score is populated from the source file, and depending on the scoring method, the range can vary. The higher the confidence, the more confidence is that the ploidy score is correct, lower range can be specified	Numeric value, range can vary depending on source.
CNV Location: in Genes from*	Selection of genes that are to be used for patient stratification based on Copy Number Variation. At least one of the below criteria must be specified.	N/A
Ad-hoc List: Gene; Pathway; Gene Set	For more information on Microarray Expression, refer to description under Microarray Expression section on page 3-17.	As in Microarray Expression section.
At Genomic Position	Specify genomic location for the variants to occur in: the format should be chr#:from-to or chr# if the entire chromosome is to be used.	chrX:13000-120000, chr7
Assembly Version	Represents GRCh version. Default selection is the last loaded reference version.	GRCh38
DNA Reference Version	Represents the Ensembl reference version for getting gene annotations. This is related to *Assembly Version* and by default displays the preferred DNA Reference Version, which is set in OHO.	Ensembl version 70

For Metadata Filter details for Microarray Expression, see Metadata Filters on page 3-30.

3.1.3.4 RNA-seq Expression

The RNA-seq Expression criteria selection interface falls under the category of Genomic criteria where you can query for patients based on results present in the W_EHA_RSLT_RNA_SEQ table and related tables in the OHO model.

As for Gene Expression, you can optionally select specimen type and anatomical site. Next, you can filter results based on the RPKM values, Raw Counts, Median length and strand.

Finally, you should specify the location for searching RNA-seq expression results which is the gene or genomic position of interest.

Figure 3–20   RNA-seq Expression

Table 3–15   RNA-seq Expression

Prompt Heading	Definition	Sample Value or Values
Specimen Type, Anatomical Site	For more information on RNA-seq Expression, refer to description under Microarray Expression section on page 3-17.	As in Microarray Expression section.
RPKM	Represents 'Reads Per Kilobaseq exon Model per million mapped reads', calculated expression intensity values in positive float or zero.	N/A
Raw Counts	Represents raw read counts in positive floating point values or a zero if unavailable.	N/A
Median Length (Normalized)	A normalized region length calculation in positive float or zero.	N/A
Strand	Strand of stored gene.	N/A

Table 3–15   (Cont.) RNA-seq Expression

Prompt Heading	Definition	Sample Value or Values
RNA-seq Location: in Genes from*	Selection of genes that are to be used for patient stratification based on RNA-seq expression. At least one of the below criteria must be specified.	N/A
Ad-hoc List: Gene; Pathway; Gene Set	For more information on RNA-seq Expression, refer to description under Gene Expression section.	As in Microarray Expression section.
For Transcript IDs	Search for Ensembl Transcript ID	N/A
At Genomic Position	Specify genomic location for the variants to occur in, the format should be chr#:from-to or chr# if entire chromosome is to be used.	chrX:13000-120000, chr7
Assembly Version	Represents GRCh version. Default selection is the last loaded reference version.	GRCh38
DNA Reference Version	Represents the Ensembl reference version for getting gene annotations. This is related to *Assembly Version* and by default displays the preferred DNA Reference Version, which is set in OHO.	Ensembl version 70

3.1.3.5 Metadata Filters

All genomic criteria screens like Sequence Variant, Copy Number Variation, Microarray Expression and RNA-seq Expression have additional filter criteria based on the metadata associated with the specimens or patients. This option is present at the bottom of each of the genomic criteria screens.

Figure 3–21   Metadata Filters

Once you expand the **Metadata Filters** option, click **Add Metadata Attribute** to open the Select Metadata Attribute dialog.

Figure 3–22   Select Metadata Attributes

Then search based on **Attribute Name**, **Scope** or **Category** to get a list of attributes associated with Metadata. Select attribute to add as Metadata Filter and assign a value for the added filer.

Table 3–16   Metadata Filters

Prompt Heading	Definition	Sample Value or Values
Attribute Name	Represents the metadata qualifier tag from W_EHA_QUALIFIER table.	N/A
Scope	Scope is retrieved from the table W_EHA_QLFR_TABLE. Based on the internal mapping scope is shown as 'Per Result' for value in table as 'W_EHA_RSLT_FILE_SPEC_QLFR' and 'Per Specimen' as scope value for value in table as 'W_EHA_RSLT_SPEC_QLFR'.	N/A
Category	Represents the metadata qualifier category tag from W_EHA_QLFR_CATEGORY table.	N/A
Value	Based on the selected metadata attribute this value datatype would change. If a numeric attribute is selected then a numeric value is given as input. Similarly, for date attribute, date would be as input and for character attribute, character would be as input.	N/A

3.2 Selecting Cohort Criteria

To select cohort criteria:

1. Navigate the **Cohort Query** tab.

2. On the left, select **Patient Information**, **Clinical Data** or **Genomic Data** to view the topics within each. For more information on the topics within each category, see Categories and Topics on page 3-1.

> **Note:** You can select criteria within any one or all the three categories.

3. Click the topic that aligns with the data you want to select.

4. Select the required data. Required fields in a topic are identified with an *asterisk*. For example, you are required to indicate one or more Diagnoses for a patient along with a date, onset, or other parameters.

   Click the magnifying glass icon to search for data for a given field.

5. In the **Insert as** group, select **inclusion** or **exclusion** to confirm whether the definition is to be included or excluded in the query.

   > **Tip:** Prepare the query definition in a structured format (line-by-line) before creating it in OHTR. Visualizing your definition may help you recognize a simple way to organize it.

6. Click **Submit**.

   To clear all the fields, click **Reset**. To close the window without making any selection, click **Cancel**.

7. The criteria you selected appears in its query statement on the right side of the page.

   You can add as many criteria as you want. The query definition expands each time you add a new statement.

3.3 Working with Inclusion and Exclusion Criteria

Once you have added several criteria statements to your query, you can:

- expand or collapse the display for the Inclusions or Exclusions criteria
- select the arrow to the left of the criteria statement to expand or collapse its detail

3.3.1 Modifying Details of an Existing Criteria Statement

You can continue to adjust the details of an existing statement using the icons on the right.

- The pencil icon displays and edits your criteria selection.
- The red X removes any particular statement.
- The green + icon adds another set of criteria from the same topic. For example, clicking the plus icon from within a **Demographics** tab prompts you to select additional Demographics criteria.

3.3.2 Logic within Inclusion and Exclusion Criteria

Within an Inclusion statement, distinct criteria statements related to the same topic is considered with an OR criteria.

For example,

INCLUDE: Diagnosis = Diabetes AND (Demographics = Male OR" Demographics = Married)

Within an Exclusion statement, distinct criteria statements related to the same topic will be considered with an AND.

For example,

EXCLUDE: Diagnosis = Diabetes OR (Demographics = Male AND Demographics = Married)

This function can be used to simplify your query structure, where you can address each data topic and category one at a time rather than having the same topic listed in both the Inclusion and Exclusion sections.

3.4 Viewing Patient Count and Patient Data

Based on the Cohort Criteria, the corresponding Patient or Subject Count is displayed. Total number of patients or subjects in the database is also displayed. This count is updated or refreshed each time you log into the system and select the Query Patients tab.

To run a query and view the counts:

1. Select at least one row of criteria defined in the **Inclusions** or **Exclusions** options.

2. Click **Run Query**. The patient count is visible in the top right.

3. To view patient data, click the patient count.
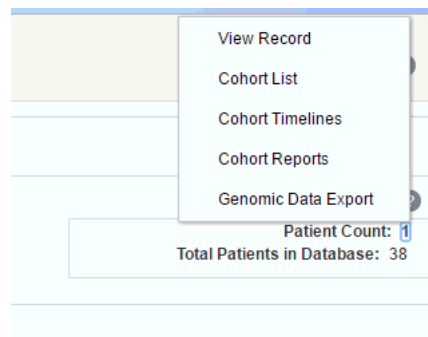
3.5 Viewing Specimen Count

To view the specimen count:

1. Click the **Options** drop-down menu in the **Cohort Query** tab.

2. Select **Show Specimen Count**.

3. The specimen count is displayed. This is the count of all the specimens in the cohort that match the selected criteria.
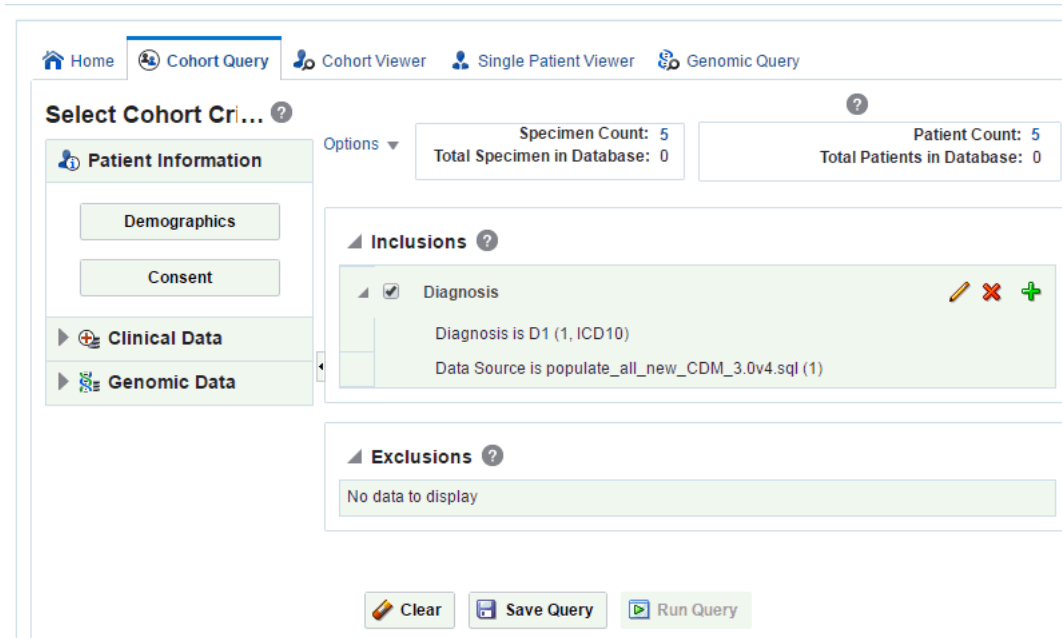
3.6 Saving a Query

When you save a query, the only information that is preserved is the criteria definition itself. Neither the patient count nor any of the underlying patient data is saved. Each time the query is opened, it must rerun to view current patient counts. Therefore, patient counts for the same query may change over time.

Each query is saved along with the context it was generated against and hence any criteria not relevant to that context is not saved.

To save a query:

1. Click the **Query Library** drop-down.

2. Click **Save Query**.
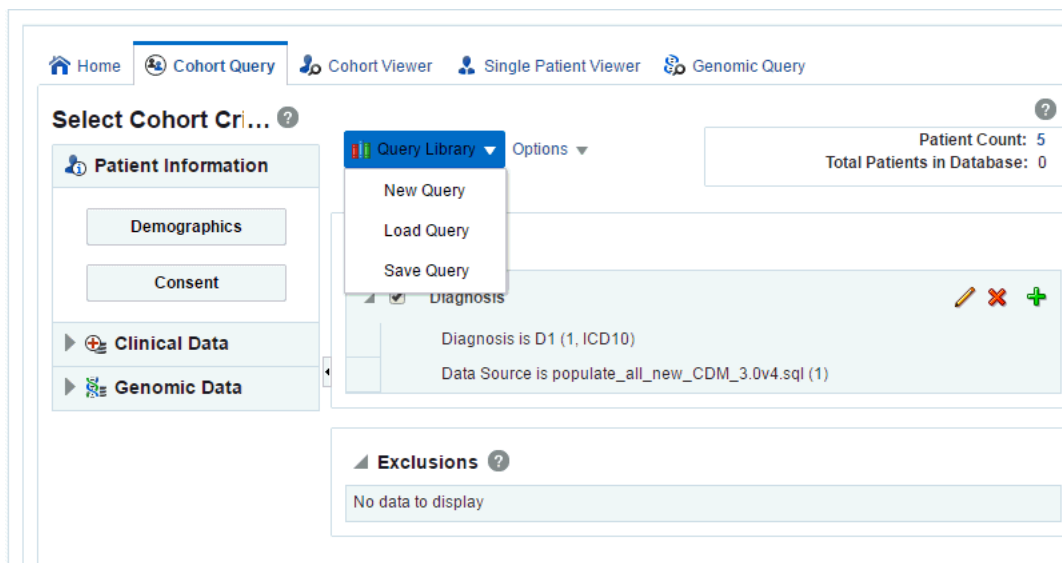
3. You can save it as a new query (with a new name), or overwrite an existing query from the library.

   If you enter a new query name and that name is already recognized in your library, the system prompts you to determine if you want to overwrite your existing query. If not, you can specify a new query name.

4. Click **Submit**. A confirmation is displayed that the query was saved.

3.7 Loading a Saved Query

To rerun a query and edit its definition, perform the following steps:

1. Click the **Query Library** drop-down.

2. Select **Load Query**.

3. Enter all or part of a **Query Name**.

4. Click **Search**.

5. Select the required query.

6. Click **Submit**. The query definition is displayed on the main page.

4

Working with the Cohort Viewer

This chapter contains the following topics:

4.1 Introduction

Cohort viewers enables you to view patients or subjects in a variety of formats. With an Oracle Healthcare Omics (OHO, formerly known as ODB) license, you can look at genomic data in a circular genome viewer (using Visquick) or export data into formats acceptable by the Integrative Genome Viewer.

4.2 Viewing a Cohort List

After the patient or subject count is displayed, you may want to review the data for these patients or subjects. You can view this data from the Cohort Queries tab as well. For details, see Section 3.4, "Viewing Patient Count and Patient Data".

You can view a cohort list for:

- a current active query

- a query from the library

- saved patient or subject list

- on an *ad-hoc* basis if you have specific patients or subjects you want to examine

4.2.1 Viewing Cohort List for Current Active Query

To view a cohort list for the query currently loaded in the Cohort Query tab:

1. Click the **Cohort Viewer** tab.

2. On the left side, check that **active query** is selected under **Patient IDs source**.

3. Click **Submit**. The cohort list and details are displayed on the right.

Use the pane on the left to select and filter the data you want to view for the cohort list. For details on filtering data, see Section 4.2.3, "Filtering Data in Cohort List".

4.2.2 Viewing Cohort List for Library Query, List or Ad-hoc

To view the cohort list for a library query or from a saved list or an ad-hoc query:

1. Navigate to the **Cohort Viewer** tab.

2. On the left side, select one of the following:

   - library query - to view the cohort list for a patient or subject query from the library

   - list - to view the cohort list for a saved patient or subject list

   - ad-hoc - to view list for specific patients or subjects you want to examine

3. Click **Submit**. The cohort list details are displayed on the right.

4.2.3 Filtering Data in Cohort List

The left hand panel contains topics that reflect the query selection criteria from the Clinical Information category. To filter data for the cohort list:

1. Click **Data** on the left hand panel to view the topics under it.

2. Select one or more check boxes for the data you want to display in your list.

3. Click **Submit**. The system reruns the query and adds relevant patient or subject data for each patient or subject listed.

A patient or subject may have multiple data for a particular data category, for example, a patient or subject can have more than one procedure or medication. Consequently, the cohort list details may display multiple rows or records for each

patient or subject. Selecting several data categories to display may significantly limit your ability to review multiple patients or subject at the same time.

4. To remove data, clear the appropriate boxes and select **Submit**.

5. To export the list to a Microsoft Excel sheet, click **Export**. To display the date in the dd/mm/yyyy format, use the formatting option in Microsoft Excel.

4.2.4 Displaying Reference Range Values

Cohort List supports displaying Reference Range values, if available, along with numeric results of Observation. To display reference range values:

1. Click **Data** on the left hand panel to show the topics within it.

2. Select **Test or Observation**.

3. Enter the value you want to search for.

4. Click **Submit**. Where available, the reference high and reference low range values are displayed along with numeric results.

4.3 Viewing Cohort Timelines

You can view patient or subject data for a small subset of patients in the form of a timeline. This process involves the following steps:

1. Selecting Patients or Subjects

2. Selecting Data

3. Displaying Patient or Subject Data

4. Aligning Data by Patient or Subject Event

5. Including Criteria Used in Query

4.3.1 Selecting Patients or Subjects

This step is similar to how you select patients or subjects for the **Cohort List**. You can select patients or subjects either from the current active query, from a query in the library, or by entering the Patient or Subject ID number.

To select patients or subjects:

1. Navigate to **Cohort Viewer** > **Cohort Timelines**.

2. Select the source of your subject ID from among the active query, library query, saved list or an ad-hoc query.

3. Click **Make initial pool**.All the subjects or patients in the selected query are added to the initial pool.

Figure 4–1 Select Patient or Subjects

4. Now select up to 20 patients or subjects and move them from the initial pool to the display list by clicking the arrow button. This selection is what the system will reference for displaying data in the timelines view.

5. Click **Submit** to view the corresponding data.

4.3.2 Selecting Data

After you have identified patients or subjects for the Display List, select the clinical information to be displayed in the viewer.

1. In the left hand panel, click **Data**. The system displays a list of data topics.

    Each data topic has a distinct color associated with it. This is so that data can be visually separated for the same patient or subject. You can change the assigned colors by using the drop-down arrow next to each color.

Figure 4–2   Selecting Data

2. Select each data topic you want to view. Then click the magnifying glass icon for each selected topic to search or specify the particular criteria. A popup search window is displayed.

3. For each topic, enter the details for the topic.

4. Click **Search**.

5. Select one or more items and use the arrow button to move them to the right hand box.

6. Click **Submit**.

7. Once you have selected and specified the data you want to view, select **Submit** at the bottom of the left hand panel. A linear view of the selected data is displayed.

4.3.3 Displaying Patient or Subject Data

The subject and patient data display is divided into two sections.

- The left side has attributes about the selected patients

- The right side shows the colored layers of the selected clinical data, in a timeline view.

To move the section divider, hover the cursor on the white vertical line. Select and drag the divider to the desired position. You can hide either section, by clicking the small left hand arrow on the right hand side of the respective area.

Click the Subject ID to navigate to the single patient viewer.

Note: If you are using Mozilla Firefox, click the patient ID. On Microsoft Internet Explorer and Google Chrome, double-click the patient ID.

- To view the data for a specific date range, use the **Time Scale** option for the date range. Enter the **From** and **to** dates and click **Update**.

  The default **Auto** option displays the data without a specific time reference.

- If an event has both Start Date and End Date values, then the timeline uses the Event Code to mark the start of an event.

- A tool tip on the timeline provides additional information about the event.

- On selecting one event in the timeline, the **Detail** table below the timeline lists all the occurrences of the event and provides additional information such as the start and the end date of the occurrences.

- For Test or Observation events with numeric result values, the **Detail** table displays the results of all the occurrences of the event. If two or more occurrences of Test or Observation event have numeric results available, then the **Graph** tab displays a line graph for those numeric result values.

- Since the Test or Observation event has only one date event, the user interface uses dummy End Date values that are 1 day later than the corresponding Start Date values to render the graph. The Detail table also displays these dummy End Date values.

- Events are not displayed on the left hand table or on the timeline in the following conditions:

  – When an event's Start Date is null.

  – When an event has only one Date attribute (for example, Observation has only Observation Date) and the attribute's value is null.

■ If an event's End Date is null, then the System Date is used for displaying the event.

4.3.4 Selecting the Timeline Mode

Use the **View same Events** option to display the events in one of the following modes:

Single Line

This is the default mode of displaying data on the Timelines tab.
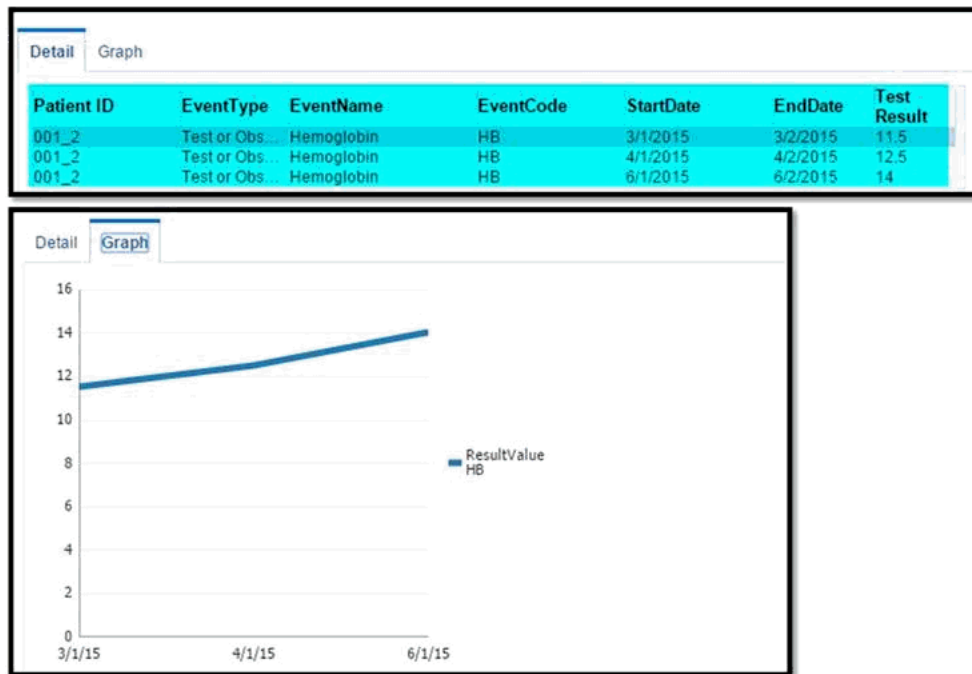
In this mode, all instances of a repeating event are displayed on the same line and separate events are displayed on separate lines.

Overlapping occurrences are displayed on separate lines in the timeline. However, the left hand table displays a single row corresponding to all occurrences of the event.

The tabular display on the left side uses alternating background color to visually delineate data of different patients or subjects. All of the rows corresponding to a patient or subject, for a particular Event Type, have one background color and the next set of rows are in a different color.

Multiple Lines

Patient ID	Event Type	Event Code	Event Name	Age in Years	Gender Name
1001	Medication	M1	Medication 1	40	Male
1001	Medication	M2	Medication 2	40	Male
1002	Medication	M1	Medication 1	35	Female
1002	Medication	M2	Medication 2	35	Female
1002	Medication	M1	Medication 1	35	Female
1003	Medication	M3	Medication 3	50	Male
1003	Medication	M3	Medication 3	50	Male
1003	Medication	M2	Medication 2	50	Male
1004	Medication	M1	Medication 1	25	Female
1001	Observation	CN	Cyanosis	40	Male
1001	Observation	HB	Hemoglobin	40	Male
1001	Observation	PRU	Proteinuria	40	Male
1001	Observation	HB	Hemoglobin	40	Male
1002	Observation	CN	Cyanosis	35	Female
1002	Observation	HB	Hemoglobin	35	Female
1002	Observation	PRU	Proteinuria	35	Female
1002	Observation	CN	Cyanosis	35	Female
1003	Observation	HB	Hemoglobin	50	Male
1003	Observation	HB	Hemoglobin	50	Male
1003	Observation	HB	Hemoglobin	50	Male
1003	Observation	PRU	Proteinuria	50	Male
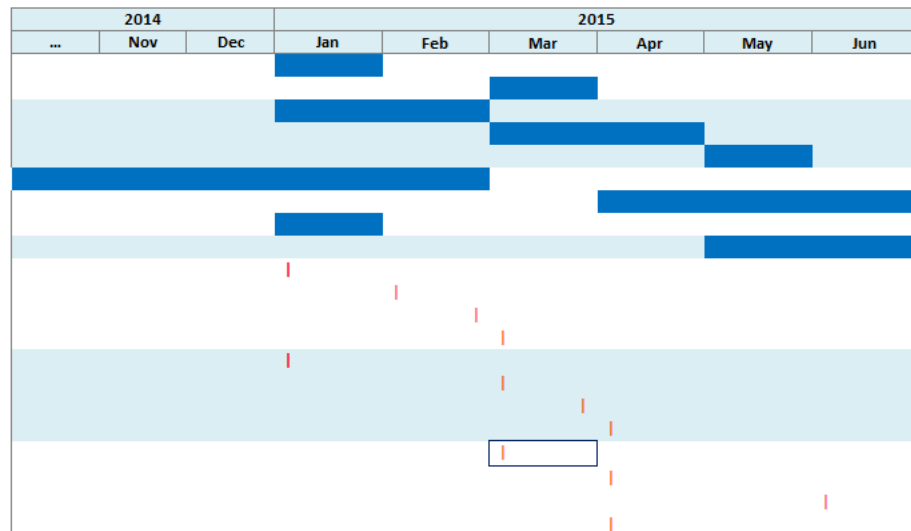
In this mode, all of the instances of events are displayed on different lines, even if the events are repeating.

Patient ID	Event Type	Event Code	Event Name	Start Date	End Date	Test Result
1003	Observation	HB	Hemoglobin	1-Mar-15	1-Mar-15	11.5
1003	Observation	HB	Hemoglobin	1-Apr-15	1-Apr-15	12.5
1003	Observation	HB	Hemoglobin	1-Jun-15	1-Jun-15	14
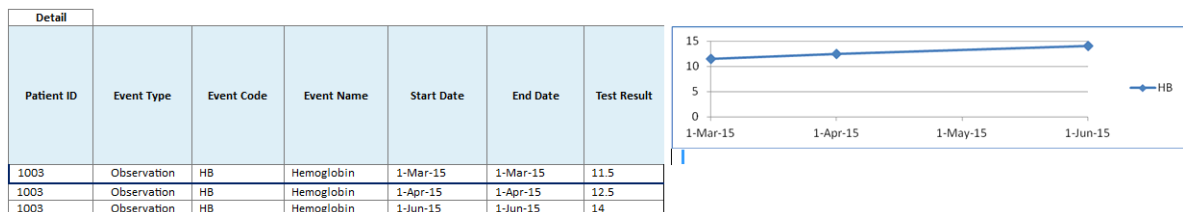
4.4 Aligning Data by Patient or Subject Event

You can designate a particular piece of clinical data to serve as the anchor, around which the remaining data is adjusted in the viewer. For example, if you choose a particular procedure as the anchor for alignment, the medications or diagnostic tests are adjusted in the display to show how long before (or after) the anchor they were administered. To do this:

1. Click **Align Data by Subject Event**. The Align Data window is displayed.

2. Enter **Subject ID**. You can also use the drop-down arrow to select a subject ID.

3. Click the magnifying glass icon next to the Align by field and select the option you want to use.

    When you select a particular diagnosis or medication code to align patients, all the event codes at different hierarchical nodes are considered for the selected event code.

4. Click **Submit**.
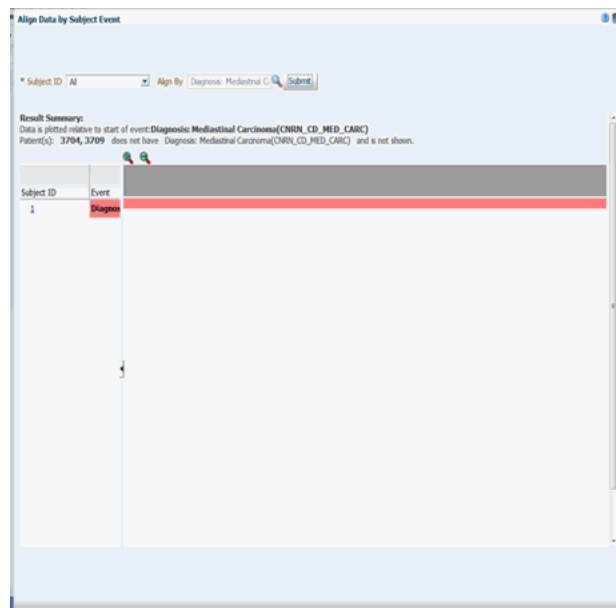
4.5 Including Criteria Used in Query

You can view the data used to define the inclusion criteria of the currently active query by selecting the **Include criteria used in query** option in the Data section of the left hand panel. The criteria defined in the cohort query is then displayed in the data section of the cohort timeline.

If the inclusion criteria of the active query is not based on any data topics supported on the cohort timeline, a message is displayed to select at least one supported data element.

> **Note:** The data topic *Clinical Encounter* can be selected only in conjunction with other data elements that are linked to Clinical Encounter, for example Diagnosis, Procedure and so on.

For example, in the Figure 4–3, **Diagnosis** is selected to be included in the criteria used in the query. After you select it, the corresponding diagnosis criteria is displayed. Click **Submit** to view the corresponding timeline events.

Figure 4–3   Include Criteria Used in Query
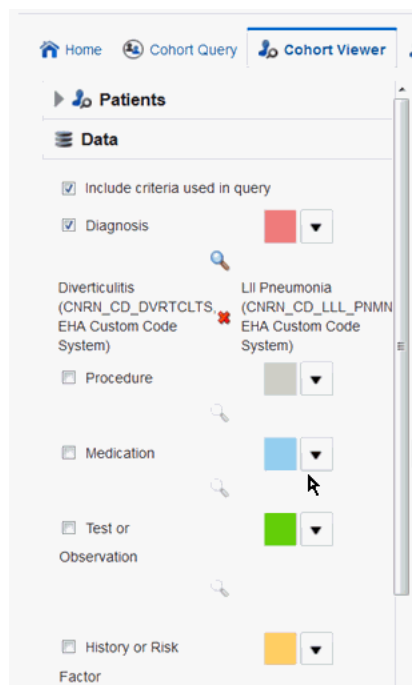
4.6 Viewing Cohort Reports

The Cohort Reports tab lets you view various cohort reports.

> **Note:**   In the Active Query mode, if no filters have been selected in Cohort Query, then the reports display data of only the first 10,000 patients or subjects. This limit is configurable and can be changed using the DEFAULT_ACTIVE_QUERY_LIMIT property in the TRC.properties file.
>
> However, irrespective of the patient or subject selection option, if the count exceeds the value specified in GENOMIC.REPORTS_MAX_ PATIENT_COUNT of TRC.properties file, a warning message is displayed. You can either continue to plot with a large number of patient or subjects (might impact performance) or change the selected cohort to a smaller number.

4.6.1 Demographic Reports

You can view demographic reports for a patient list. The queries for these reports can be based on patient list in the cohort query, the library, patient list, ad-hoc query or omics. Perform the following steps to view demographic reports:

1. Navigate to **Cohort Viewer** > **Cohort Reports** tab.

2. Expand the **Patients** node on the left.

Figure 4–4   Patient Node
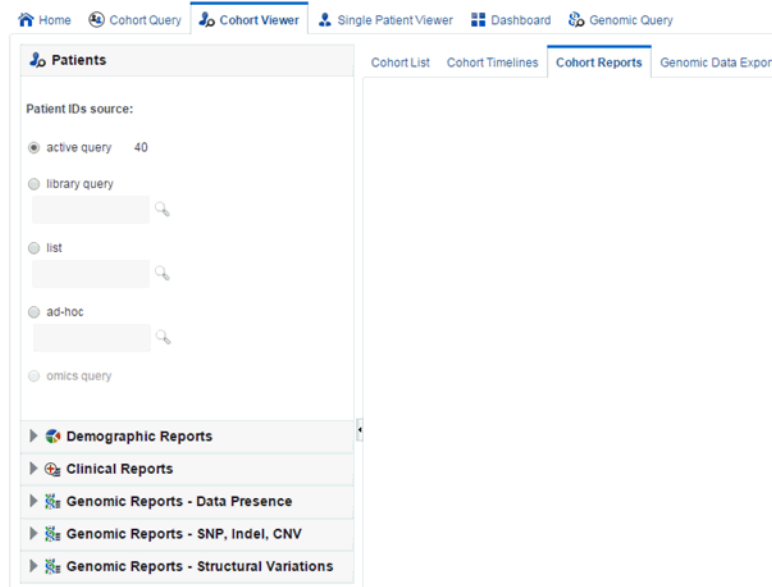
3. Select the **Patient IDs Source**.

4. Expand the **Demographic Reports** tab.

Figure 4–5   Select Demographic
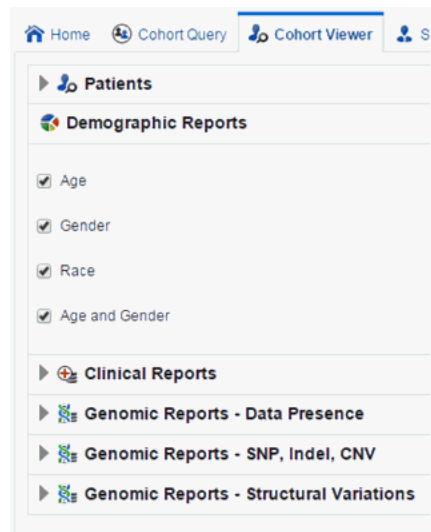
5. Select the check boxes for the data you want displayed in your list.

6. Select **Submit**.

7. The system reruns the query and adds relevant patient or subject data for each patient or subject listed. Depending on you selection, reports age, gender, race and/or age and gender will be displayed. The data will be displayed as pie charts or bar graphs depending on your settings.

Figure 4–6   Demographics Report
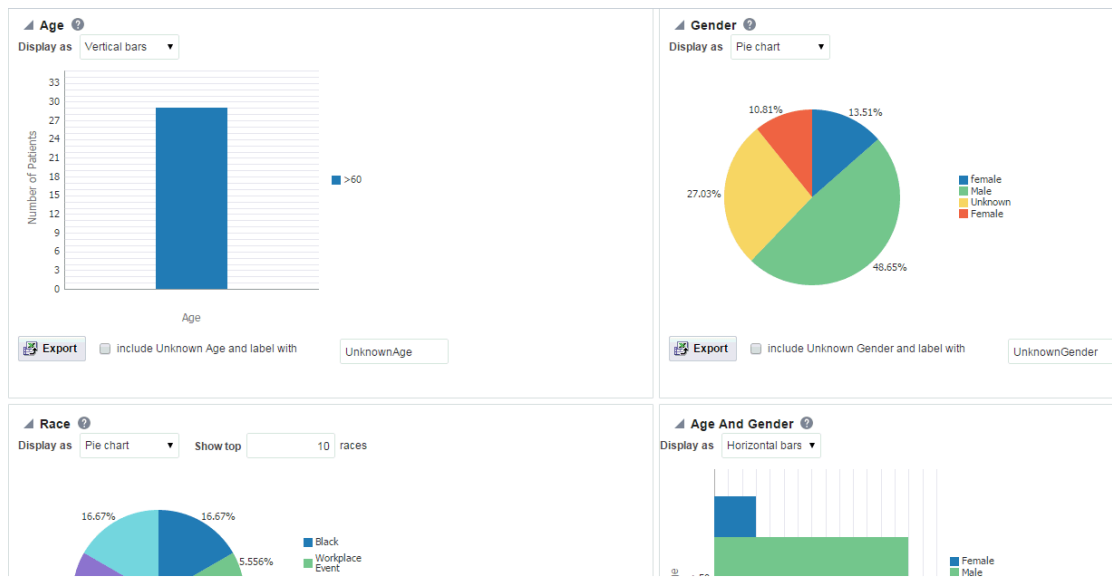
4.6.1.1 Handling Unknown Data

If any missing or unknown data is present, a check box is provided at the bottom of each graph to include missing or unknown data as shown in Figure 4–6.

For the Age, Gender and Race graphs, if the check box is selected, the graph is refreshed with the unknown data. Here unknown data refers to age, gender or race details that are not defined for a particular patient.

For the Age and Gender graph, the unknown age for different genders is combined into one group and the known age is in other group. Both groups are shown in the graph.

4.6.2 Clinical Reports

You can view clinical reports for a patient list. The queries for these reports can be based on patient list in the cohort query, the library, patient list, ad-hoc query or omics. Perform the following steps to view a clinical reports:

1. Navigate to **Cohort Viewer** > **Cohort Reports** tab.

2. Expand the **Patients** node on the left.

Figure 4–7   Patient ID Source
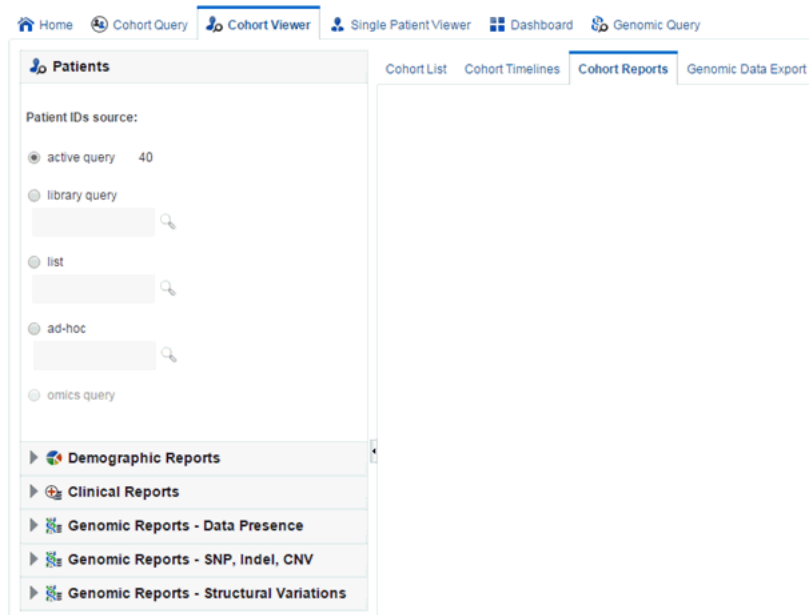
3. Select the **Patient IDs Source**.

4. Expand the **Clinical Reports** tab.

5. Select the **Specimen Available** check box.

6. Select **Submit**.

7. The system reruns the query and adds relevant patient or subject data for each patient or subject listed. Depending on you selection, reports age, gender, race and (or) age and gender will be displayed. The data will be displayed as pie charts or bar graphs depending on your settings.

Figure 4–8   Clinical Report
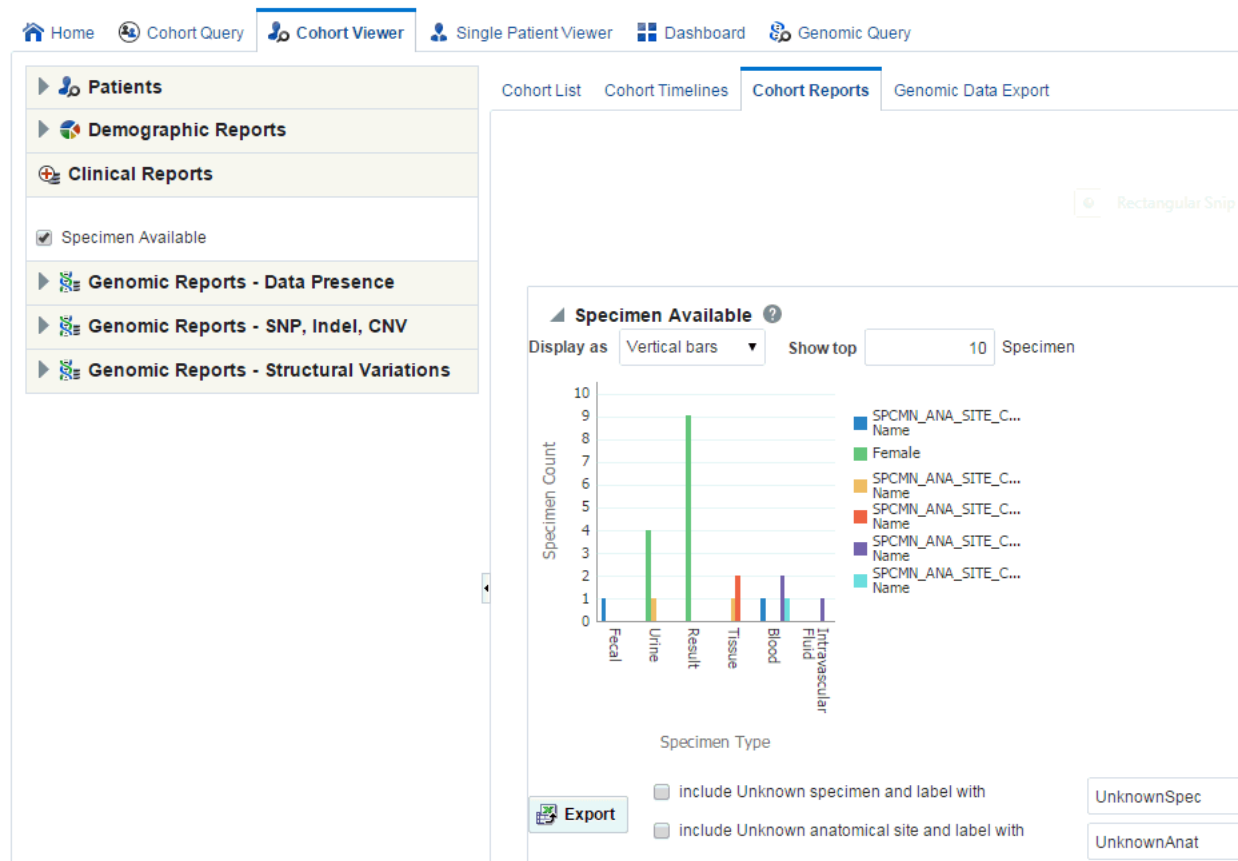
4.6.2.1 Handling Unknown Data

If any missing or unknown data is present, a check box is provided at the bottom of each graph to include missing or unknown data as shown in Figure 4–8.

If **Include Unknown Specimen and Label With** or **Include Unknown Anatomical site and Label With** are selected, the graph is refreshed with the unknown data. Here unknown data refers to the specimen and anatomical site data that is not defined for a particular patient.

4.6.3 Genomic Reports

4.6.3.1 Data Presence

The Genomic Reports - Data presence lets you view a breakup of the patient cohort as follows:

- Patients in the cohort with specimens collected that have accessible genomic data

- Patients in the cohort with specimens collected but no genomic data

- Patients in the cohort with no specimens collected

You can view the Data Presence plot as Pie Chart, Horizontal Bars, Vertical Bars, and Table. The report also provides an option to view the distribution of present genomic data along the following available genomic data types:

- Sequencing

- RNA-seq

- Gene Expression

- Copy Number Variation

Perform the following steps to view a Genomic data presence report:

1. Navigate to **Cohort Viewer** > **Cohort Reports** tab.

2. Expand the **Patients** node on the left and select the source of patient or subject identifiers.

3. Expand the **Genomic Reports - Data Presence** node in the left panel.

4. Select the **Genomic Data Presence** check box. An optional **Specimen Type** input field appears to allow for filtering the initial cohort on specimen type.

5. Click **Submit** to generate the report.

Figure 4–9 Genomic Data Presence
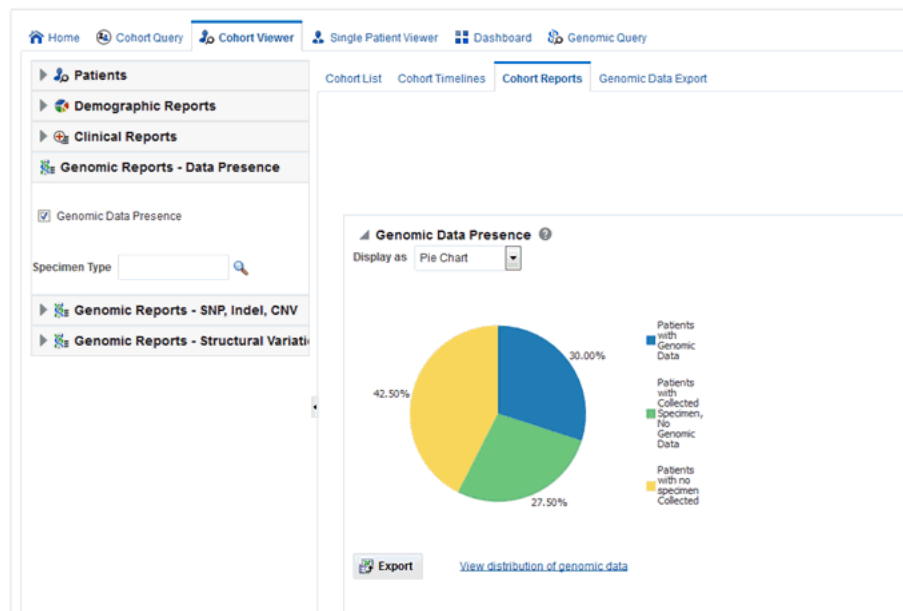
6. To change the display type after generating the report, select display option from the **Display as** drop-down list.

Figure 4–10   Data Presence - Table Display

7. To view a breakup of present genomic data by data type, click **View distribution of genomic data**.

Figure 4–11   Distribution of Genomic Data
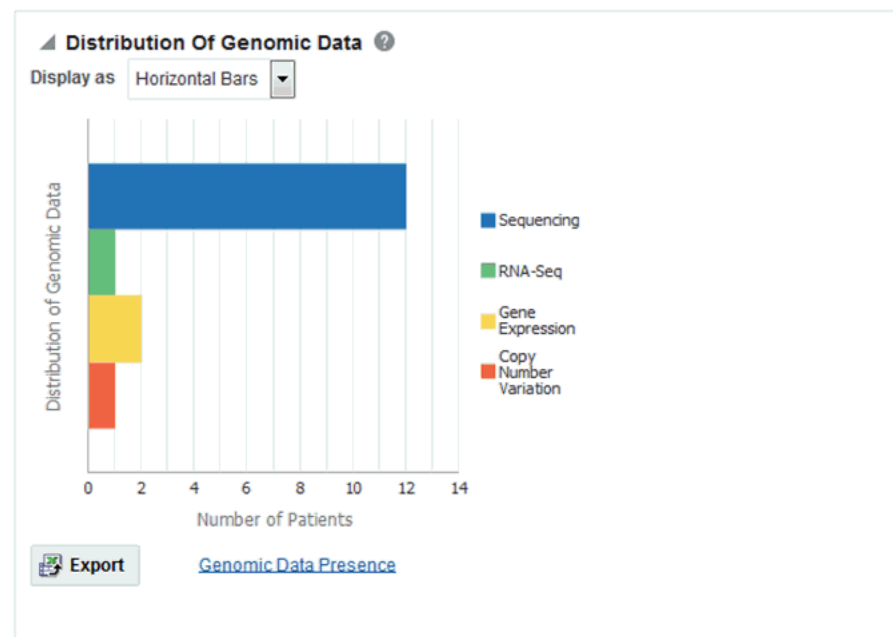
8. To export the graphical plot type displays as a PNG image, and the table type displays as a Excel spreadsheet, click **Export**.

4.6.3.2 SNP, Indel and CNV

4.6.3.2.1   Gene Level Reports - Mutated Gene Frequency and Gene Expression

Genomic Reports - SNP, Indel and CNV display the available SNP Indel genomic reports based on the selected cohort of patients or subjects, if in subject context.

To view a SNP, Indel and CNV report:

1. Navigate to **Cohort Viewer** > **Cohort Reports** tab.

2. Expand the **Patients** node on the left and select the source of patient or subject identifiers.

3. Navigate to **Genomic Reports - SNP, Indel, CNV** and select **Gene Level Reports.**

4. Select **Mutated Gene Frequency and Gene Expression**. You can also add additional parameters such Specimen Type, Anatomical Site, Assembly Version which will only consider results linked to the selected categories.

   Selecting **Include All Specimen** includes specimens without genomic data whereas not selecting this returns a result with only genomic data.

5. Click **Submit**. A histogram report appears with the percentage of samples for the relevant cohort, which have sequence variants information within the selected genes.

   You can display the results as horizontal bars, vertical bars, or as a table. To export the bar results into PDF or the table results into Excel, click **Export**.

Note: When opening the Excel file, you may receive a warning from Excel stating that the file is in a different format than specified by the file extension. This warning can be safely ignored. For more information, see http://docs.oracle.com/cd/E23943_01/web.1111/b31973/af_table.htm#autoId34.

Figure 4–12   Show the Sequence Variants
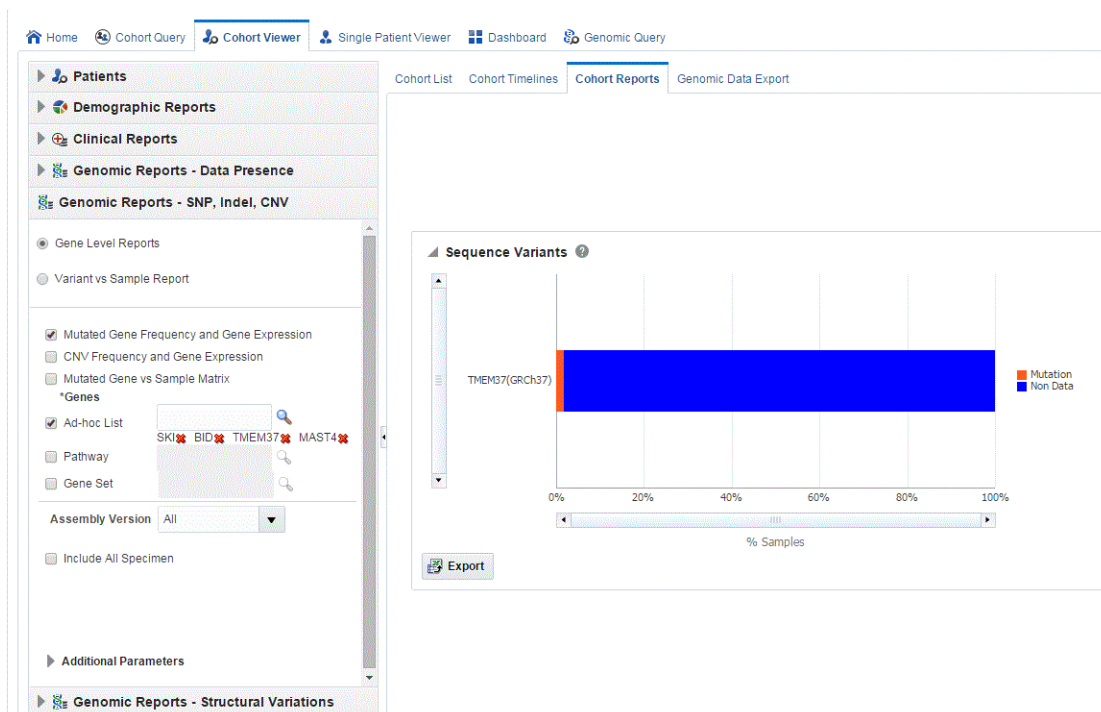
6. Clicking the histogram for the different selected genes displays a popup to either get the Gene Expression plot or the Variant Viewer to display details of the variant.

Figure 4–13   Popup
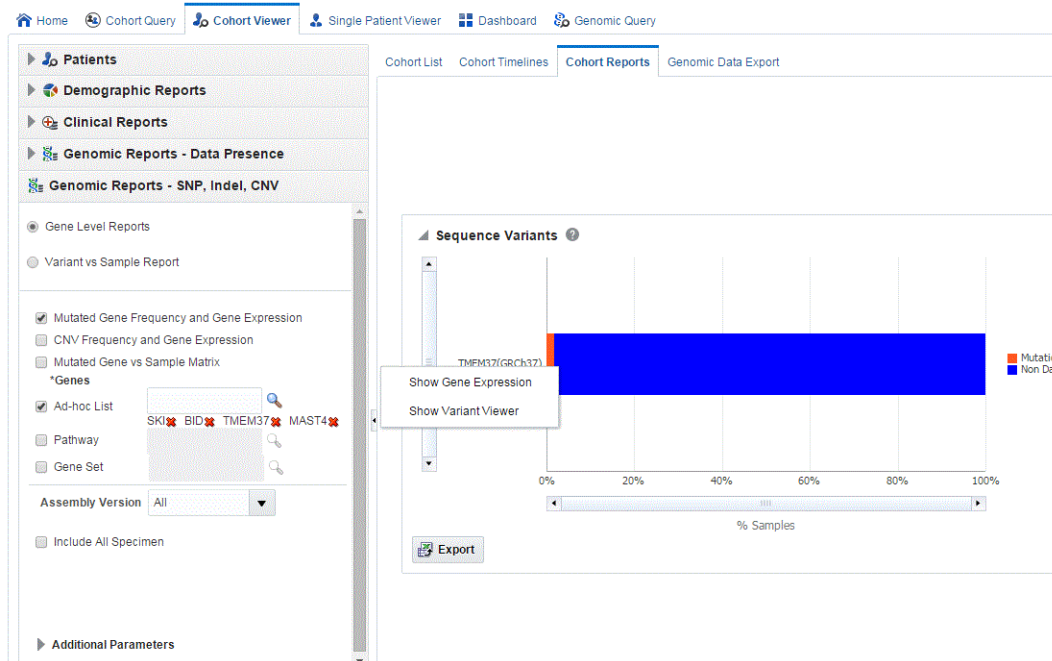
7. To view the Gene Expression details for Single Channel or Dual Channel, click **Show Gene Expression** in the popup. In the Single Channel Expression, you must provide hybridization details.

8. Selecting **Show Variant Viewer** displays necessary information about all variants present in that gene in the samples belonging to the selected cohort of subjects or patients.

The Mutation plot displays the total number of specimen (on *y*-axis) having a particular mutation (on *y*-axis) falling in the selected gene for belonging to a specific reference. The *x*-axis contain the range of gene region including the flanking region.

The CDS plot displays bars of CDS regions for each of the Ensembl Transcripts belonging to the selected gene of a selected reference version. Each row in the CDS plot represents one Transcript ID of the selected gene and reference version.

Details of each variant like Variant name, Variant Type, Replace Tag, Variant Effect if present, Location, Disease if loaded, Histology, Site Total number of samples, Patients are displayed. The results are grouped in descending order of specimen count, such that each row contains unique variants.

Variant Name	Variant Type	Replace Tag	Variant Effect	Location	Disease	Histology	Site	Samples
COSM112142,rs...	insertion	-/GTGTGC	in-frame insertion	chr2:120194651		carcinoma, aden...	large_intestine, c...	4
COSM112142,rs...	insertion	-/GTGTGC	in-frame insertion	chr2:120194651		carcinoma, serou...	ovary, NS	4
COSM112142,rs...	insertion	-/GTGTGC	in-frame insertion	chr2:120194651		carcinoma, aden...	large_intestine, r...	4
COSM112142,rs...	insertion	-/GTGTGC	in-frame insertion	chr2:120194651		carcinoma, aden...	large_intestine, c...	4

9. You can drill down to each mutation by clicking **Number of samples** present in the samples column of the table. A table and a pie chart are displayed below the table. The table contains details such Sample ID, Study (only in subject context), Anatomical Site and Patient or Subject ID, which contains the selected mutation, number of samples which do not contain this mutation (labeled as Other) from the group of samples having mutations for the selected gene and selected reference version.
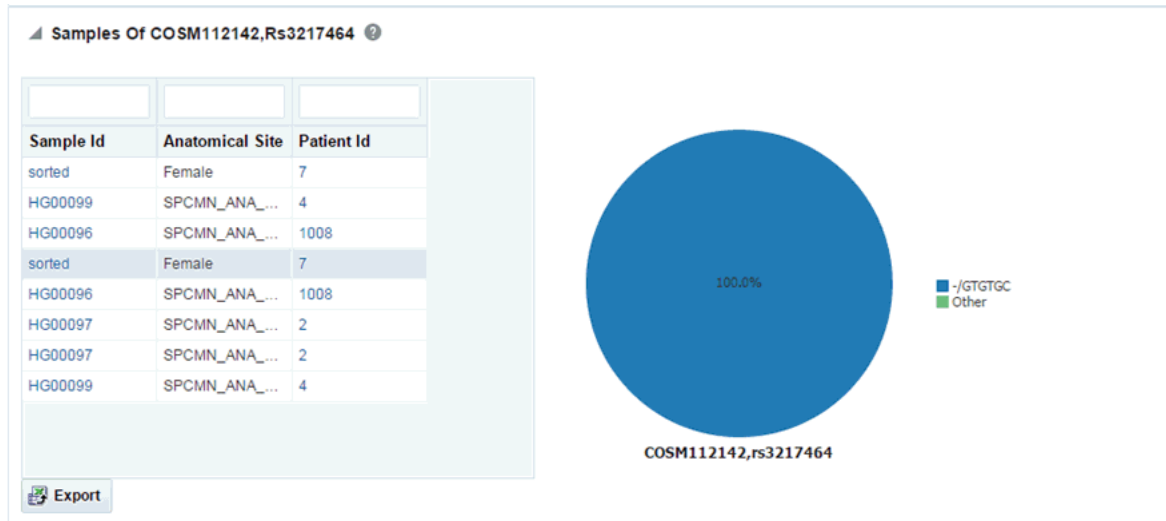
4.6.3.2.2 Copy Number Variation Frequency and Gene Expression

To view CNV Frequency and Gene Expression reports:

1. Navigate to **Cohort Viewer** > **Cohort Reports** tab.

2. Expand the **Patients** node on the left and select the source of patient or subject identifiers.

3. Navigate to **Genomic Reports - SNP, Indel, CNV** and select **Gene Level Reports.**

4. Select **CNV Frequency and Gene Expression**. You can also add additional parameters such Specimen Type, Anatomical Site, Assembly Version, which will only consider results linked to the selected categories.

   Selecting **Include All Specimen** includes specimens without genomic data whereas not selecting this returns a result with only genomic data.

Figure 4–14   Select Source of Patient or Subject Identifiers
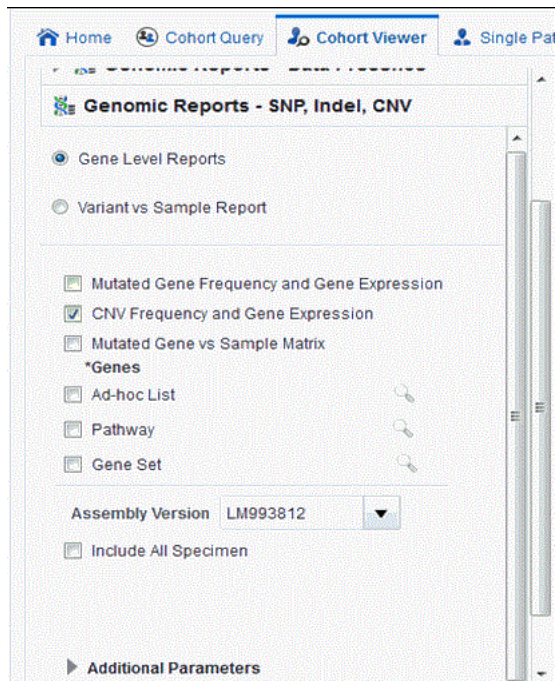
5. Click **Submit**. A histogram report displays the percentage of samples for the relevant cohort, which have copy number variants information within the selected genes.

    You can display the results as horizontal bars, vertical bars, or as a table. To export the bar results into or table results into Excel by clicking **Export**.

    > **Note:**   When opening the Excel file, you may receive a warning from Excel stating that the file is in a different format than specified by the file extension. This warning can be safely ignored. For more information, refer to http://docs.oracle.com/cd/E23943_01/web.1111/b31973/af_table.htm#autoId34.

Figure 4–15   Show the Copy Number Variations
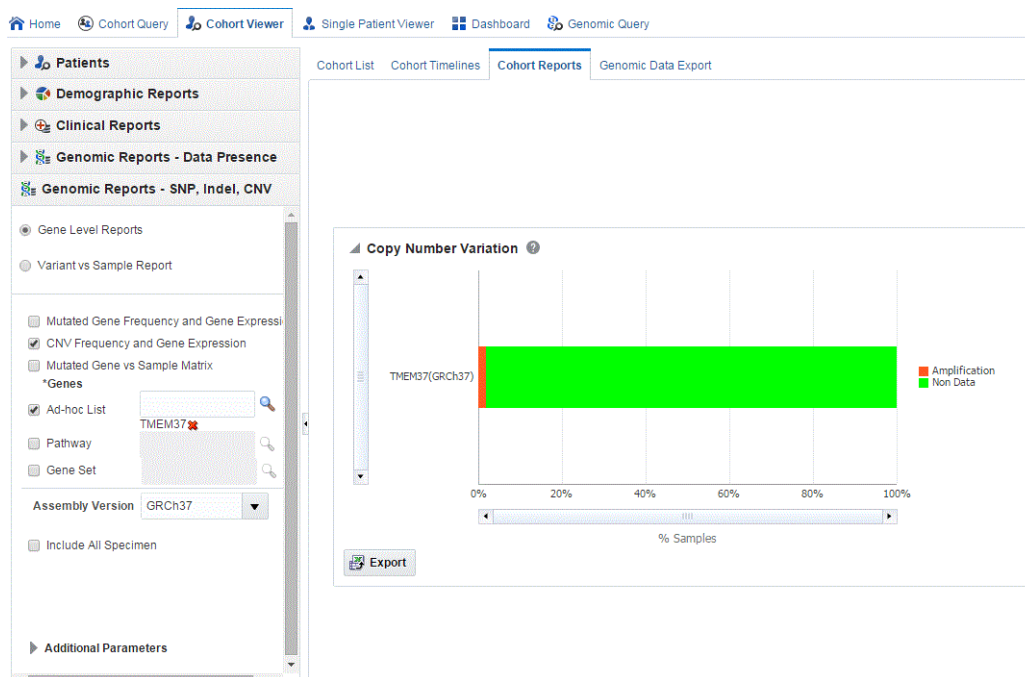
6. Clicking the histogram displays a popup is to either get the Gene Expression plot or the CNV Viewer to display details of the variant.
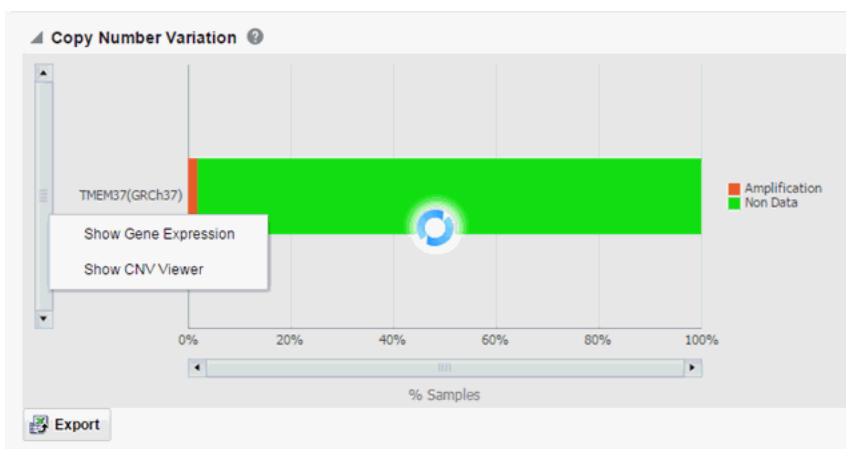
7. To display the Gene Expression details for Single Channel or Dual Channel, select **Show Gene Expression**. In the Single Channel Expression, you must provide hybridization details.

8. Clicking **Show CNV Viewer** displays necessary information about all CNV variants that are present in that gene in the samples belonging to the selected cohort of subjects or patients.

   The CNV plot displays each specimen (on *y*-axis) and its values falling in the selected gene for belonging to a specific reference.The *x*-axis plots the range of gene region. The plot is color coded with CNV with Gain (CNV > 0.2) is Red, Normal (-0.2<CNV<0.2) in grey and Loss (CNV < -2) in Green.
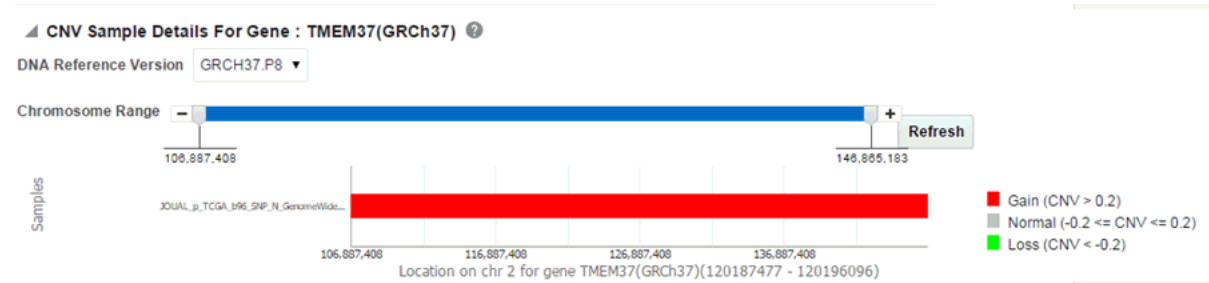
The CNV table provides details of CNV data such as Samples, File Type, Chromosome, Anatomical Site, Start Position, End Position, CNV Value and Patient or Subject ID.

Sample Id	File type	Anatomical Site	Chromosome	Startposition	Endposition	CNV Value	Patient Id
JOUAL_p_TCGA...	CNV .seg file	SPCMN_ANA_SI...	2	106887408	146865183	0.205	6

4.6.3.2.3 Mutated Gene vs Sample Matrix

The Mutated Gene vs Sample Matrix plot displays a high-level pictorial view of the presence of specific variants in genes of interest across various specimens of patients or subjects in a cohort. You can see if a particular gene from a specimen has selected mutations or no-mutations or no-data loaded. The same plot lets you view CNV data on the genes for the specimens.

The specimens are grouped together for each patient or subject and are ordered based on their collection date.

You can also view *patient_id:specimen_number(specimen_vendor_number):collection_dt:no of variants* in the tooltip for each data point.

Variant selection is based on variant impact, implying that the depiction of mutations on genes for each specimen is based on the variants having selected variant impacts. For example, if you select *frameshift* mutation for EGFR, then the query searches for variants causing frameshift impact for EGFR gene and reports as *mutation* if it finds any such variants. If there is only non-variant information, then it reports as non-mutation in the plot. If there is no data, then it shows no-data in the plot.

Following are the definitions of various types of data in the plot.

Data	Definition
Mutation	If any of the variants are present in the selected gene of a specimen, then it is reported as mutation.
No Mutation	If none of the variants are present in the selected gene of a specimen and there is non-variant information available for that gene, then it is reported as No Mutation.
No data	If there are no variants and also no non-variant information and there is no-call data for the selected gene of a specimen or there is no-data loaded for this specimen, then it will be reported as no data.
Amplification	If there is CNV data, which has seg_mean value more than zero, then it is reported as amplification.

Data	Definition
Deletion	If there is CNV data which has seg_mean value less than zero, then it is reported as deletion.

To view this plot:

1. Navigate to **Cohort Viewer** > **Cohort Reports** tab.

2. Expand the **Patients** node on the left and select the source of patient or subject identifiers.

3. Navigate to **Genomic Reports - SNP, Indel, CNV** and select **Gene Level Reports.**

4. Select **Mutated Gene vs Sample Matrix**.

5. Select genes either by ad-hoc list, pathway or gene set. Data is filtered using the Variant Impact or the cosmic mutation selected.You can also add additional parameters such Specimen Type, Anatomical Site, Assembly Version, which will only consider results linked to the selected categories.

    Selecting **Include All Specimen** includes specimens without genomic data whereas not selecting this returns a result with only genomic data.

6. Click **Submit**. A report displays the percentage and details of samples for the relevant cohort which have sequence variants and copy number variants information within the selected genes.

Figure 4–16   Mutated Gene vs Sample Matrix Report

Each bar represents a specimen, which is grouped patient-wise and in the order of the specimen collection date. The Sequence Variants and CNV information for each specimen is displayed. The percentage of specimens with mutation on the gene is also mentioned. Export functionality is provided at each specimen level.
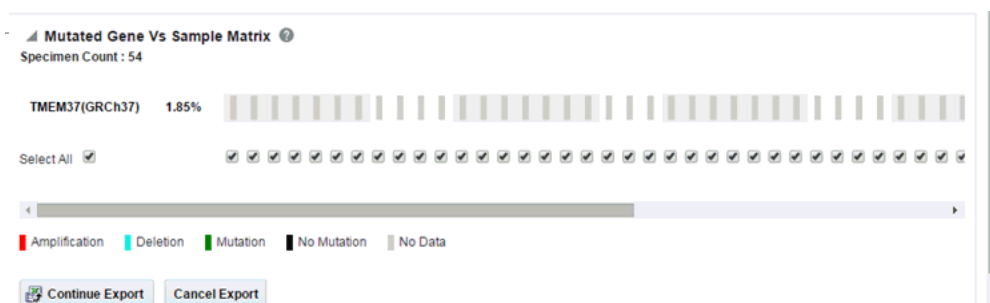
> **Note:** When opening the Excel file, you may receive a warning from Excel stating that the file is in a different format than specified by the file extension. This warning can be safely ignored. For more information, refer to http://docs.oracle.com/cd/E23943_01/web.1111/b31973/af_table.htm#autoId34.

There is a limit to the number of specimens that can be seen in the gene matrix plot. This is an application level parameter called 'MAX_SPEC_REPORT' that is defined in the TRC.properties file. The default value of this parameter is 1000. If the number of specimens in the cohort used for this analysis is greater than the specified value, then the following warning message is displayed:

If you continue from this warning message, the report displays the summary statistics instead of the matrix plot. You may have to decrease the cohort size based on some criteria like specimen type or anatomical site or otherwise use cohort query to view the matrix plot. Alternatively, you can also change the default value of MAX_SPEC_REPORT parameter to a desired value and rebuild the plot. However, rebuilding maybe affect the performance of the plot generation.

The summary report only displays the percentage of specimens with different categories of data as shown in the image above.

4.6.3.2.4 Variant vs Sample Reports

This option is displayed on the Genomic Reports - SNP Indel genomic reports available, based on the selected cohort of patients or subjects, if in subject context. It provides a high level pictorial view of the presence of specific variants of interest across various specimens of patients or subjects in a cohort.

The specimens are grouped together for each patient or subject and are ordered based on their collection date. You can also view *patient_id:specimen_number(specimen_vendor_number):collection_dt* in the tooltip for each data point.

To generate the plot:

1. Navigate to **Cohort Viewer** > **Cohort Reports** tab.

2. Expand the **Patients** node on the left and select the source of patient or subject identifiers.

3. Navigate to **Genomic Reports - SNP, Indel, CNV** and select **Variant vs Sample Report**.
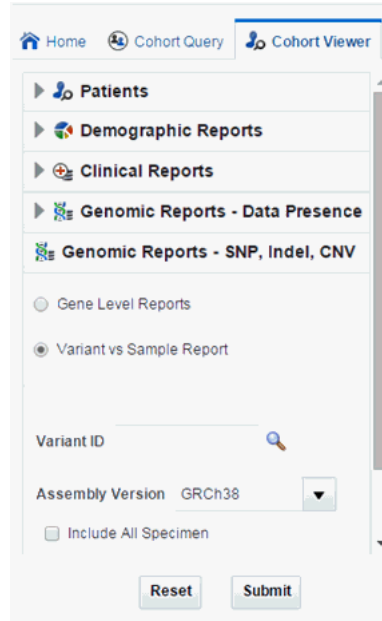
4. Select **Variant ID**.

   You can also add additional parameters such Specimen Type, Anatomical Site, Assembly Version, which will only consider results linked to the selected categories.

   Selecting **Include All Specimen** includes specimens without genomic data whereas not selecting this returns a result with only genomic data.

5. Click **Submit**. A report displays the percentage and details of samples for the relevant cohort, which have mutation within the selected genes.

Figure 4–17   Variant vs Sample Report
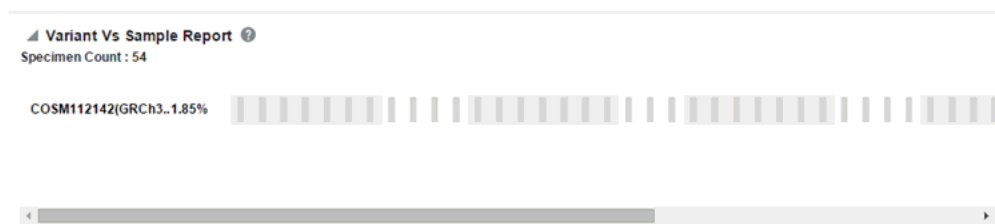
Each bar represents a specimen, which is grouped together patient wise and in the order of the specimen collection date. The mutation information for each specimen is displayed. Each row represents one variant and the label represents reference_ id, assembly version and the variant replace_tag value to maintain the uniqueness of the variant. The percentage of specimens for each mutation is also mentioned.
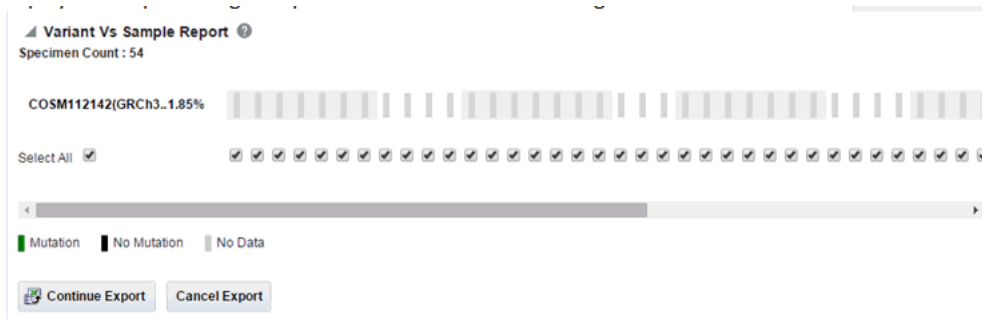
> **Note:** When opening the Excel file, you may receive a warning from Excel stating that the file is in a different format than specified by the file extension. This warning can be safely ignored. For more information, refer to `http://docs.oracle.com/cd/E23943_01/web.1111/b31973/af_table.htm#autoId34`.

There is a limit on the number of specimens that can be seen in matrix plot. This is an application level parameter called 'MAX_SPEC_REPORT' that is defined in the TRC.properties file. The default value of this parameter is 1000. If the number of specimens in the cohort used for this analysis is greater that the specified value then the following warning message is shown below:

If you continue from this warning message, the report will display the summary statistics instead of the matrix plot. You may have to decrease cohort size based on some criteria like specimen type or anatomical site or otherwise using cohort query to view the matrix plot. Alternatively, you can also change the default value of MAX_SPEC_REPORT parameter to a desired value and rebuild the plot but the performance of the plot generation may be effected. The summary report will only display the percentage of specimens with different categories of data as shown below.

4.6.3.3 Structural Variations

4.6.3.3.1 Structural Variations in Genes

To view the structural variations in genes:

1. Navigate to **Cohort Viewer** > **Cohort Reports** tab.

2. Expand the **Patients** node on the left and select the source of patient or subject identifiers.

3. Navigate to **Genomic Reports - Structural Variations** and select **Structural Variations in Genes**.

   You can also add additional parameters such as specimen type, anatomical site, DNA version, which will only consider results linked to the selected categories.

Figure 4–18   Select Source of Patient or Subject Identifiers

4. Click **Submit**. A histogram report shows the occurrence of structural variations involving genes sorted from the gene involved in most SVs for a given cohort, and subsequently display genes with decreasing frequency of Structural Variations. By default, the top 10 genes are shown and you can increase or decrease the number of genes displayed.

5. You can display the results as horizontal bars, vertical bars, or as a table. To export the bar results into PDF or table results into Excel, click **Export**.

> **Note:** When opening the Excel file, you may receive a warning from Excel stating that the file is in a different format than specified by the file extension. This warning can be safely ignored. For more information, refer to http://docs.oracle.com/cd/E23943_01/web.1111/b31973/af_table.htm#autoId34.

Figure 4–19 Show the Structural Variations

4.6.3.3.2 Structural Variations in Gene Pairs

Structural Variations in Gene Pairs is one of the SV genomic reports available to you based on the selected cohort of patients or subjects, if it is in the subject context.

1. Navigate to **Cohort Viewer** > **Cohort Reports** tab.

2. Expand the **Patients** node on the left and select the source of patient or subject identifiers.

3. Navigate to **Genomic Reports - Structural Variations** and select **Structural Variations in Gene Pairs**.

You can also add additional parameters such as specimen type, anatomical site, DNA version, which will only consider results linked to the selected categories.

Figure 4–20   Select Source of Patient or Subject Identifiers
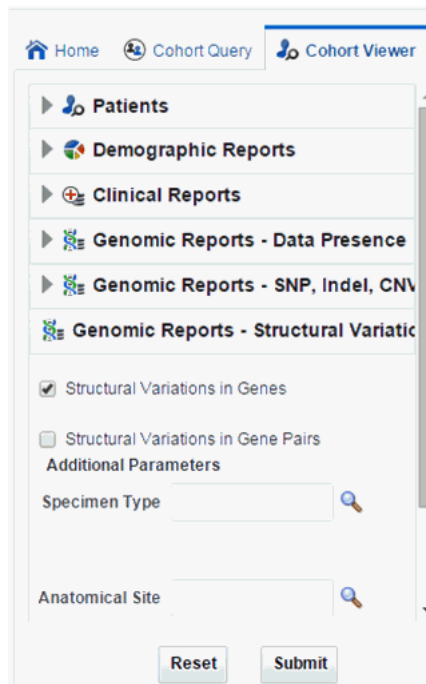
4. Click **Submit**. A histogram report displays the frequency of occurrence of structural variations in a cohort among gene pairs. The histogram is automatically sorted from the gene pair with most SVs as per cohort and the incidence decreases or at best stays the same for each subsequent gene pair. By default, top 10 gene pairs are shown and you can change the default number of bars shown.

5. You can display the results as horizontal bars, vertical bars, or as a table. To export the bar results into PDF or the table results into Excel, click **Export**.

> **Note:**   When opening the Excel file, you may receive a warning from Excel stating that the file is in a different format than specified by the file extension. This warning can be safely ignored. For more information, refer to http://docs.oracle.com/cd/E23943_01/web.1111/b31973/af_table.htm#autoId34.

Figure 4–21   Show the Structural Variations
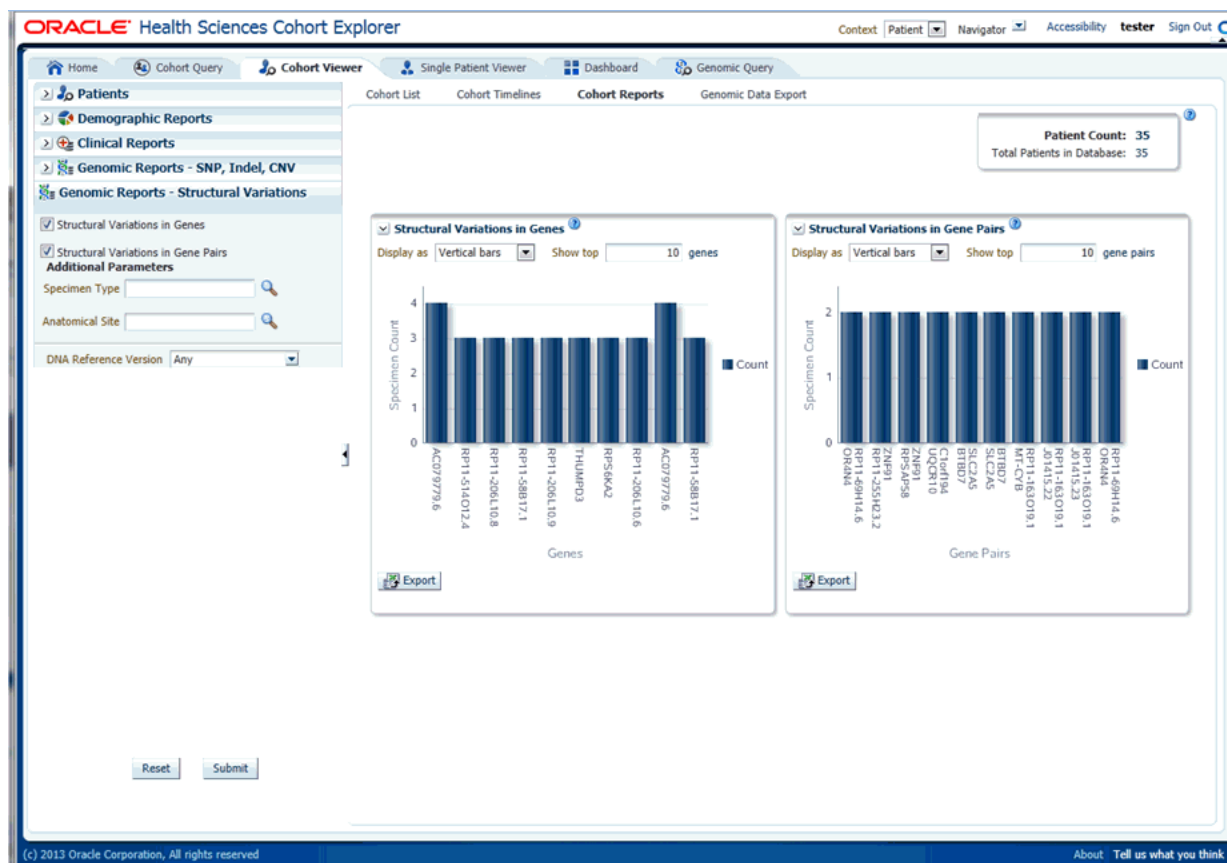
4.7 Exporting Genomic Data

The **Genomic Data Export** page is used to export the genomic data for patients or subjects filtered based on Study, Specimen type and Anatomical Site in a specific file format.

Currently, exporting variation data from sequencing platform, single and double channel gene expression and Copy Number Variation data in VCF, SEG, RES and GCT file formats are supported. These formats are supported by the IGV browser.

Figure 4–22   Genomic Data Export

4.7.1 Selecting Patients or Subjects

You can download data for patients or subjects already selected in an active query or from a query library or from ad-hoc list of patients ID. There is no upper limit on the number of patients or subjects to be selected, however the performance slows down as more and more patients are selected.

4.7.2 Selecting Results to Export

Figure 4–23   Selecting Results to Export

After selecting Patients, select the **Assembly Version**, **Specimen Type**, and **Anatomical Site**. These selection criteria will help you filter out patients based on the requirements. **Specimen Type** and **Anatomical Site** also have multiselect options. Currently only one version of data can be exported at a time.

4.7.3 Selecting Location

Figure 4–24   Selecting Location

You can download genomic data for either a list of genes or pathway or a gene set for a defined region in chromosome. You can also export the genomic data for a specific chromosome region and also the complete genomic data for the patient using the **All Data** option.

On Exadata, the code takes advantage of chromosome based partitioned data for VCF and SEG export. This enables more accurate results to be exported, including intergenic result. On non-Exadata systems, only the results that lie within any gene boundary are exported.

4.7.3.1 In Genes From

You can select genes from one or more of the provided three options.

■	Using ad-hoc List, you can select one or more genes.

■	Using Pathway, you can select one or more pathways which in turn will get the list of genes associated with the selected pathway internally for querying.

■	With Gene Set, you can use the user-defined collection of genes.

The genomic data to be downloaded is based on the above selected genes.

4.7.3.2 At Genomic Position

You can alternatively download genomic data using the genomic co-ordinates. You specify the chromosome region in a standard format for the Variation and CNV data to be exported. The Gene Expression - RES and Gene Expression - GCT download option would be disabled for genomic region criteria. You can specify a complete chromosome or a part of chromosome as criteria. Currently, only one chromosome region at a time is implemented for search.

The following chromosome region formats are supported.

■	CHR15:10000-200000: Considers region between 10000 to 200000 in chromosome 15.

■	CHR15:1,200,000+5000 - Considers 5000 bases upstream from 1,200,000 position in chromosome 15.

■	CHR15 - Considers whole of the chromosome 15.

■	CHR15:1000 - Considers 1000th nucleotide position of chromosome 15.

4.7.3.3 All Data

The genomic location selection is **All Data** option, which is only available for Schedule download and not for immediate download. With this option you can download all the data available for the specimens belonging to the selected patients or subjects falling under the selected criteria.

4.7.4 Selecting File Type

This panel lists out four file type options to export. You can select all four options at a time. For Genomic Region criteria, the Gene Expression – RES, and Gene Expression Dual Channel options are disabled.

Click **Submit**. The data is generated and a link is provided in the bottom panel separately for each result type.

Figure 4–25   File Type to Export options

4.7.4.1 Mutation - VCF

This option exports the sequencing variation data for the selected patients or subjects for either the selected genes, pathway, geneset or for a given chromosome region as selected in the previous option. VCF supports multiple specimens' data in a single file.

The metadata header gives the following information that differs based on the search criteria:

1. ##fileformat=VCFv4.1

2. ##fileDate: Date and time of the VCF file generated.

3. ##source=Oracle Healthcare Omics (OHO, formerly known as Omics Data Bank)

4. ##Total Number of patients included in this VCF file

5. ##Total Number of samples included in this VCF file

6. 7.##INFO=<ID=NS, Number=1, Type=Integer, Description=Number of Samples With Data>

7. ##FORMAT=<ID=GT,Number=1,Type=String,Description=Genotype>

8. ##FORMAT=<ID=GQ, Number=1, Type=Integer, Description=Genotype Quality>

9. ##FORMAT=<ID=GQVAF, Number=2, Type=Integer, Description=Genotype_quality_X>

10. ##FORMAT=<ID=DP, Number=1, Type=Integer, Description=Read Depth>

11. ##FORMAT=<ID=AD,Number=.,Type=Integer,Description=Allelic depths for the ref and alt alleles in the order listed >

12. ##FORMAT=<ID=HQ, Number=2, Type=Integer, Description=Haplotype Quality>

13. ##FORMAT=<ID=BQ,Number=.,Type=Integer,Description=Average base quality >

14. ##FORMAT=<ID=MQ,Number=.,Type=Integer,Description=Average mapping quality >

15. ##FORMAT=<ID=SS,Number=1,Type=Integer,Description=Variant status relative to non-adjacent Normal,0=wildtype,1=germline,2=somatic,3=LOH,4=post-transcriptional modification,5=unknown>

16. ##FORMAT=<ID=SSC,Number=1,Type=Integer,Description=Somatic Score>

The following data types are imported to VCF file:

Data Type	Description
CHROM	chromosome
POS	position of the variation
ID	dbSNP ID or COSMIC ID associated with a variant
REF	reference allele
ALT	variant alleles
QUAL	not populated. Will have '.' specified in this column.
FILTER	is populated as PASS
INFO	Not populated. Will have '.' specified in this column.
FORMAT:GT	genotypic data for each specimen.
FORMAT:GQ	genotype quality. If not value available in DB, then '.' is specified in the file.
FORMAT:GQX	mapped to GENOTYPE_QUALITY_X column
FORMAT:DP	this stores the TotalReadCount for a specific variant
FORMAT:AD	this stores the reference read count and Allele read count for a specific variant.
FORMAT:HQ	not populated as of now. Will have '.' specified in this column.
FORMAT:FT	this stores GENOTYPE_FILTER column value
FORMAT:BQ	stores the RMS base quality
FORMAT:MQ	stores the RMS mapping quality
FORMAT:SS	stores the somatic status
FORMAT:SSC	stores the somatic status score value
Flex field format	If any custom formats are available, they are also included in the export.

1000 Genomes VCF 4.1 conventions are followed while exporting variation data, however certain datatypes, which are non-standard, like BQ and MQ, may differ in convention for some customers since there is no standard way to represent them.

4.7.4.1.1 Handling Non-variant and No-call Data

If NON_VARIANT and (or) NOCALL records exist for any given position, the zygosity is checked to determine if the format information from these tables is used.

> **Note:** For *het-ref* or *half* zygosity values, these other format fields are compared with the existing SEQUENCING information. This information is then used with zygosity to create the format string.

The NON_VARIANT data allows for GQ, GQX, MQ, BQ and the first reference read count of AD. The NOCALL data allows for all format fields to be compared. Both NON_VARIANT and NOCALL do not support exporting flex fields. The GT value of the format string reflects the stored zygosity as follows:

Zygosity	FORMAT string GT:GQ:GQX:BQ:MQ:AD:DP
het-ref	1/0:99:98:38:45:20:10,10
Half	1/.:99:98:34,34:45,45:20:10,5
Het-alt	1/2:99:98:43,44:56,67:20:0,10,10
Hom	1/1:99:98:34,34:45,45:20:0,19

If there are no result records for any specimen, the export displays "." with no other information for the format.

4.7.4.1.2 Handling Ambiguous Sequencing Data in Export

There could be cases where users reload genetic information multiple times for the same specimen. This may create ambiguous values for the different fields that exist in the VCF export file. The export code deals with such ambiguous numerical values that represent the quality (that is, GQ, GQX, AD, BQ, MQ). This code now computes minimum values and ensure that the value of least confidence is reported. There could be more complex cases, for instance, if there are 2 different alleles for the same position belonging to the same specimen, or variants with same position for same specimen with different zygosity. The export code uses MIN functions on all values including all the text fields. This allows for VCF export to create a valid file that can be loaded into genome browsers.

Alternatively, you can choose not to consider data from a specific specimen or a specific file using following methods:

- Using DELETE_FLG - A user may load results for a specimen more than once that can completely contradict previous results. Users can set the DELETE_FLG as 'Y' on W_EHA_RSLT_SPECIMEN and (or) W_EHA_SPEC_PATIENT or W_EHA_SPEC_SUBJECT to have previous loads excluded, and then reload the correct result files. When the user now exports the data, only the latest loaded specimen data is considered for export.

- Using FILE_URI - Oracle recommends using this method since you need not reload the data again as opposed to the above method. When there are multiple files loaded with contradicting data for the same specimen, user can set some files as obsolete by changing the W_EHA_FILE_LOAD.FILE_WID column. For example, if you have loaded the same specimen data 3 times and would like to consider the latest file loaded for export, then you must first identify the latest FILE_WID from W_EHA_FILE_LOAD table. Then change the FILE_WID of two old files in W_EHA_FILE_LOAD table to the latest FILE_WID. Now, all the three records belonging to the three file loads contain same FILE_WID, which represents the latest file load and only the latest file export data is exported.

Representing AD Values

Allele depth values represented under the AD datatype are in the order of the alleles represented in the GT. Refer to the following table with examples:

ALT	FORMAT	SAMPLE1	
G,C,T	GT:AD	1/2:0,4,6	0 represents reference_read_count
			4 represents allele_read_count of 'G'
			6 represents allele_read_count of 'C'
G,T	GT:AD	2/2:0,4	0 represents reference_read_count
			4 represents allele_read_count of 'T'
G,T	GT:AD	1/0:10,5	10 represents reference_read_count
			5 represents allele_read_count of 'G'

4.7.4.2 Copy Number Variation - SEG

The copy number variation data is exported in SEG format. Currently, CNV data from any array based system like Affymetrix Genome Wide SNP 6 array whose data is in SEG format while loading in OHO is supported. The main requirement for exporting CNV data is to have the SEG_MEAN value in the CNV table of OHO.

For exporting data that is not loaded from SEG files, for example, data from CGI CNV files or any other source of CNV data, users have to create their own loader. The loader is expected to calculate the SEG_MEAN value since this value is most important for export.

1. ID: specimen ID of the reported CNV segment

2. chrom: chromosome name

3. loc.start: start position of the CNV segment

4. loc.end: end position of the CNV segment

5. num.mark: for array based CNV data, this stores the number of probes details

6. seg.mean: this stores the segment mean value from SEG_MEAN column in CNV table.

4.7.4.3 Gene Expression - RES

RES is one of the gene expression formats supported by IGV browser. Currently, only microarray gene expression data is exported to this format. Following data types are imported to RES format:

1. Description: hugo name of a specific probe

2. Accession: probe ID

3. Intensity: intensity value of the associated probe

4. Call: call of the associated probe

4.7.4.4 Gene Expression Dual Channel - GCT

GCT is one of the gene expression formats supported by IGV browser. Currently, only AgilentG4502A platform microarray gene expression data is exported to this format. Following data types are imported to GCT format:

1. Description: Gene symbol of a specific probe

2. Accession: probe ID

3. Intensity: intensity value of the associated probe

4. Call: call of the associated probe

> **Note:** The GCT file takes its gene symbol for the probe from the 2-channel composite element of ADF file. This is input into the ADF composite table in OHO. This value may not match with HUGO name in certain cases as OHTR associates 2-channel records in the result table that has partial (which includes a flanking region set by the user) genomic coordinate. The coordinate overlaps between composite elements and gene segments in the reference. This may also result in some cases in more than one unique gene in the reference mapping to a gene composite element.

4.7.4.5 Export Options

You can export data in the following two ways:

- Select option to download last loaded file(s)
- Immediately, which is the default option
- Schedule

Figure 4–26   Export Options - Schedule Mode

The **Immediately** option gives you the file link on the same screen and you can click to download it immediately. The link provided has a specific naming convention: *<file type>_OHO_<date:MM-DD-YYYY>_<time:HH24*-MI-SS>.<file_type_extension>*. For example, RES_OHO_09-14-2014_04-26.res. A short description of the file stating data type and advise on the expected count of features is displayed below the created link.

The **Schedule** option runs the process as a job. You can track the status of the job from the **Home** > **Jobs** tab. This option is best suited for exporting large data set like **All Data**, whole chromosome variants and so on. For schedule option, you must provide a job name and description. Then click **Submit** to start the process. For more details, see Working with Jobs on page 2-9.

There is a possibility of replicate and duplicate data in the database. This could be due to loading multiple files belonging to the same specimen_number. This can happen if the same library is sequenced multiple times or the data is reanalyzed, for example, the reads were realigned using the new reference version and hence new VCF or gVCF files are created for same sample. In this scenario, you can use the option to export VCF data only from last loaded files. For example if variation data has been loaded for a specimen in Jan 2015, Mar 2015 and July 2015, then using this option you can export data from the file loaded in July 2015 and it would not consider variants from the file loaded in Jan and Mar 2015.

> **Note:** The Schedule Jobs option uses an asynchronous approach to store the file in DBFS. As an alternative to downloading the file using the link in the Oracle Healthcare Translational Research Jobs page, there are other ways to access DBFS. From a Linux OS, you can mount DBFS using *dbfs_client* application and then browse the directories. Windows OS does not support the FUSE interface and cannot mount DBFS directly. However, there is a *dbfs_client* application for Windows that can execute commands to access DBFS. The Windows version of *dbfs_client* lets you use the command line to execute normal directory commands. You can list the DBFS directories as well as copy data from DBFS to the local drive. The *dbfs_client* application is part of the standard Oracle client software.
>
> For more information about using *dbfs_client*, see http://docs.oracle.com/cd/E11882_01/appdev.112/e18294/adlob_client.htm#ADLOB0006.

5

Single Patient or Subject Viewer

This chapter describes the single patient or subject viewer of OHTR. It contains the following topics:

- Section 5.1, "Viewing Records"
- Section 5.2, "Circular Genomics Viewer"

5.1 Viewing Records

The **View Record** tab helps you focus on the discreet medical history or genomic history of a particular patient. You can drill into the data details of a particular patient, and locate the attribute or data element that is pivotal as selection criteria for your cohort.

To view a patient record:

1. Enter the patient ID.

2. Select **Source**. By default, **All records** is selected, which lets search through all the patients available in the CDM database.

3. Select the clinical and genomic data you want to view for the patient.

4. Click **Submit**. The available patient information is displayed on the right.

Figure 5–1   View Record

5. To print the current screen, click **Print** at the bottom of the screen.

6. To save the data in an Excel file, click **Export** at the bottom of the relevant section.

A patient or subject may have significant volumes of information and you can search for a particular attribute. Clicking the header of any column will sort the data within it.

1. Enter a value in the blank field at the top of any column.

2. Press Enter. The rows are filtered based on the attribute you have specified.

> **Note:** The attributes that display for the various categories align with the data terms and definitions that are described in the **Cohort Query** tab, where you select criteria for selecting patients or subjects.

5.1.1 Navigating Through Selected Patients or Subjects

If you have a specified patient or subject count in the single patient or subject screen on the Cohort Query or Cohort List screen, you can navigate through all the selected patients or subjects clinical and genomic history by using the Previous and Next buttons.

Under the **Patient** field, the current Patient's ordinal number (ordinal position of the Patient record within the selection) and the total number of available   Patients or Subjects (in the Cohort) will be displayed. For example, in the following figure it is (*1 of 8*) patients.

Figure 5–2   All Records and Cohort Options
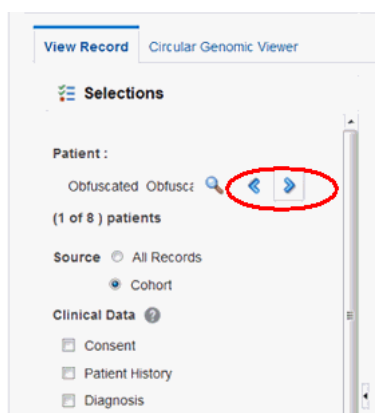
If you have navigated from the Cohort Timeline screen into the single patient or subject screen, then the total number of available patient or subject cohort will be the count of patients or subjects selected in the initial pool.

5.1.2 Genomic Data

Genomic data is displayed in four sections as follows:

5.1.2.1 Specimen with Genomic Results

Table 5–1   Specimens with Genomic Results

Column Heading	Definition	Sample Value or Values
Specimen Id	Specimen belonging to the selected patient or subject	HG00096
Specimen Vendor Id	Specimen Vendor for that specimen	Vcf
Version Label	Represents Assembly Version (DNA reference version against which this data was loaded)	GRCh37(V68)
Sequence Variants Results	Whether the sample has sequence variants results	Yes / No
Copy Number Variation Results	Whether the sample has copy number variants results	Yes / No
Single Channel Microarray Results	Whether the sample has single channel results	Yes / No
Dual Channel Microarray Results	Whether the sample has dual channel results	Yes / No
Rna-Seq Expression Results	Whether the sample has rna sequencing results	Yes / No

5.1.2.2 Derived Files

Table 5–2   Derived Files

Column Heading	Definition	Sample Value or Values
Filename	Filename including path of the genomic file stored including path	C:/John_specimen01.vcf
File Size in MB	Size of the File in MB	Numeric, positive integer
File URI	URI of the file	File://trc/abc.maf
Alternate Filepath	The FTP path of the file	
File Type, Version	Type of file and Version	Variant Call Format, 4.1

Table 5–2   (Cont.) Derived Files

Column Heading	Definition	Sample Value or Values
Result Type	Type of result data in the file	Sequencing, Copy Number Variation, Gene Expression (2-channel or single channel)
Alignment Version (DNA Reference Version)	Represents Assembly Version (DNA reference version against which this data was loaded)	GRCh37(v68)
Total Number of Specimen in File	Total number of specimen that the file contains where not all specimen belong to the selected patient	Numeric, positive integer
Last Updated	When record was last updated	19-Mar-2012

5.1.2.3 File Lineage

> **Note:**   If you have appropriate permissions, and if files are present in the middle tier accessible location, the path listed in the **File Name** fields for Genomic data are enabled to let click and download the files directly from OHTR.

Table 5–3   File Lineage

Column Heading	Definition	Sample Value or Values
Parent Filename	Parent Filename including path of the genomic file stored including path	C:/John_specimen01.BAM
File Size in MB	Size of the File in MB	Numeric
File URI	URI of the file	File://trc/abc.maf
Alternate Filepath	The FTP path of the file	
File Type, Version	Type of file and Version	Binary Alignment Map, 1.0
Alignment Version (DNA Reference Version)	Represents Assembly Version (DNA reference version against which this data was loaded)	GrCH37
Last Updated	When record was last updated	19-Mar-2012
Derived Child Files: File - Specimen Id, Vendor Id	Information about Derived Results files that have their lineage based on the particular Low Level file.	C:/John_specimen01.vcf - HG00096_1,HarvardLab1

5.1.2.4 Variants Found
Selecting this option displays the different variants available for a patient. These variants are grouped and displayed in a hierarchical structure with the count of the variants displayed for each type of variants.

Figure 5–3   Variants Found
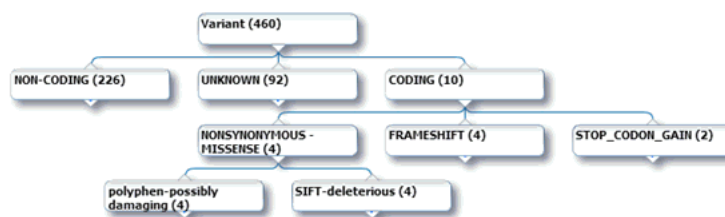
On selecting any of the nodes the details of the variants are displayed in a table as follows:

Table 5–4   Variant Details

Column Heading	Definition	Sample Value or Values
Chromosome	Chromosome location of the variant	1
Position	The position of the variant within the chromosome	Numeric value
Reference Allele	The reference allele of the variant	C
Alternate Allele	The alternate allele of the variant	G
Gene	The gene containing the variant	BID
Transcript	The transcript name	ENSTXXX
Variant Name	The reference id of the variant	rs111
Variant Type	The type of variant	Substitution
Variant Status	Status of the variant	Known
Protein Name		
SIFT Impact	The SIFT impact of the variant	Intolerant
PolyPhen Impact	The polyphen Impact of the variant	Damaging
Drug	The related drug	Clofazimine
Associated Disease	The disease associated to the variant	Anaemia
Histology		
Site		
Specimen Id	The specimen containing the variant	HG00096
Alignment Version	Alignment version of a variant	GRCh37

5.1.3 Selecting Assembly Version

By default, the last loaded assembly is displayed but this selection can be modified. The Genomic data is filtered out with this assembly version instead of the DNA reference version. One Assembly version can belong to multiple DNA reference versions.

5.1.4 Dalliance Browser

The Dalliance browser is a third party tool that displays a graphical representation of the variant, chromosome or gene range. The variant detail table comprises of hyperlinks for variant reference ID, gene name and chromosome position.

1. In the patient details section, scroll down to **Variants**.

2. The variant detail table comprises of hyperlinks for variant reference ID, gene name and chromosome position.

3. Click any of these hyperlinks to navigate to the Dalliance browser where you can see the particular VCF tracks around a specific variant, a specific gene, or in an explicitly specified chromosomal range.
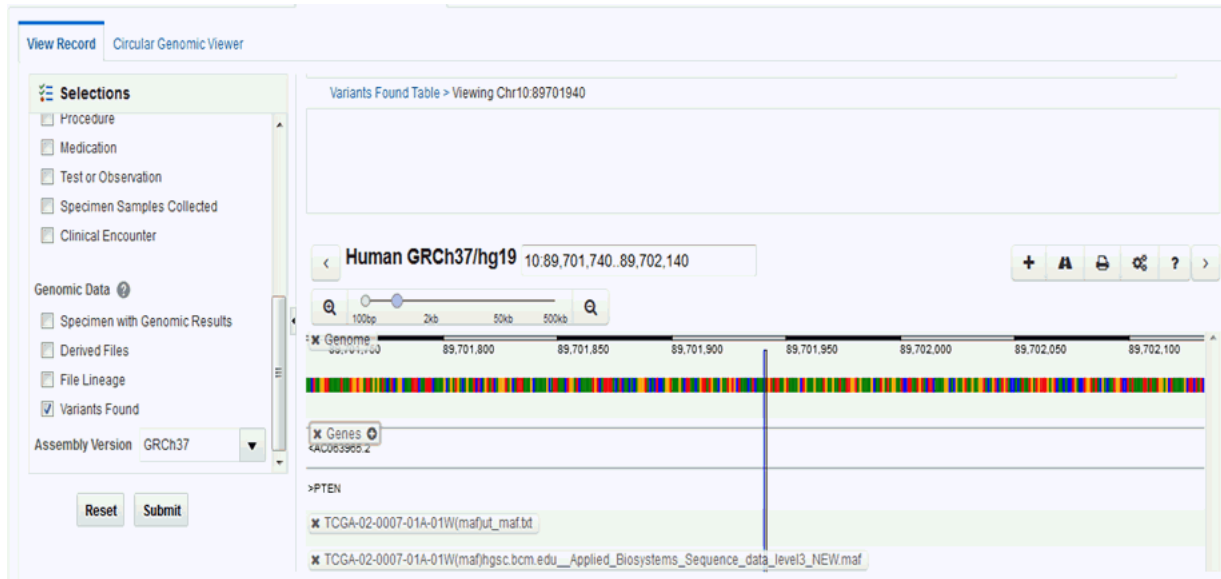
If a file is loaded multiple times for a particular specimen, then the following warning message is displayed. It lists all the multiple files names.

⚠ **Warning**
Multiple file loads have been found for TCGA-02-0007-01A-01W(maf):
(1) hgsc.bcm.edu__Applied_Biosystems_Sequence_data_level3_NEW.maf

Following are some configurations required to plot the gene track in the Dalliance browser. You can create your own DAS server and the corresponding entries should be added in the TRC.properties file. Following is an example of the required entries, to look up the Authority, University of California Santa Cruz (UCSC) name for the alignment and the public URLs for the Sequence and Genes tracks (reference tracks):

```
DALLIANCE.AUTHORITY_37=GRCh
DALLIANCE.UCSC_NAME_37=hg19
DALLIANCE.SEQ_URL_HG18=http://www.derkholm.net:8080/das/hg18comp/
DALLIANCE.SEQ_URL_HG19=http://www.derkholm.net:8080/das/hg19comp/
DALLIANCE.GENES_URL_HG18=http://www. derkholm.net:8080/das/hsa_54_36p/
DALLIANCE.GENES_URL_HG19=http://www. derkholm.net:8080/das/hsa_59_37d/
```

While integrating Dalliance with OHTR, the server `http://www. derkholm.net:8080/das/` has been used. However, this has been shut down.

These can be manually customized in TRC.properties. To add any new alignments, add all the above code to TRC.properties. Make sure that you research the reference for where they can be found. It is possible to use downloaded files instead of public DAS server(s) for reference, but the client should host these files on a web server.
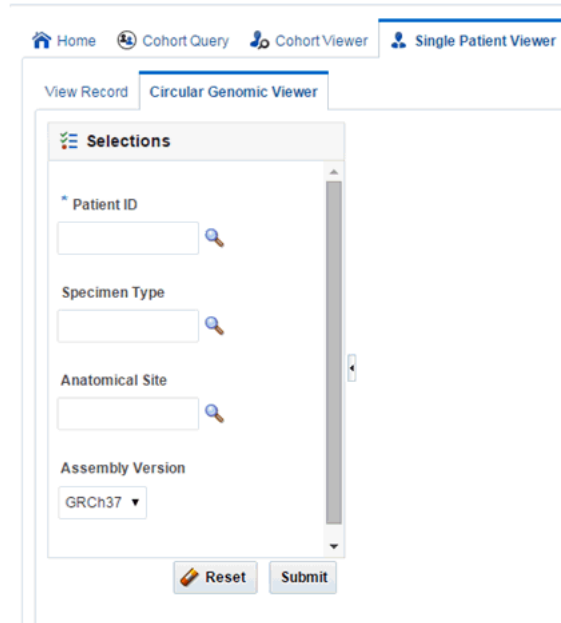
5.2 Circular Genomics Viewer

Circular genomic data viewer provides an interface for you to visualize the genomic data which includes variation, micro array expression, copy number variation, dual channel expression and RNA sequencing. The system uses the VisQuick tool, which is a Javascript library built specifically for genomic data visualization.

To view circular genomic data:

1. Select Patient ID or Subject ID. If the Patient or Subject ID has been selected in the View Record screen, it will be inherited here.

Figure 5–4   Selection Criteria

2. To add filter criteria, select **Specimen Type and Anatomical Site**. The DNA Reference Version selected is used to filter out the results and determine the cytoband to be used while rendering the circular genomic plot for any of the five data types.

3. Click **Submit**. Based on the filter criteria, the matching specimen in any of the five result types are displayed.

Figure 5–5   Specimen for Result Types

4. Select only one specimens of different result types to plot the graph. By default, the cytoband of chromosomes is also plotted which is the outer most ring of the circular plot.

5. Click **Plot Graph**. The circular genomic data graph is displayed.

Figure 5–6   Circular Genomic Plot
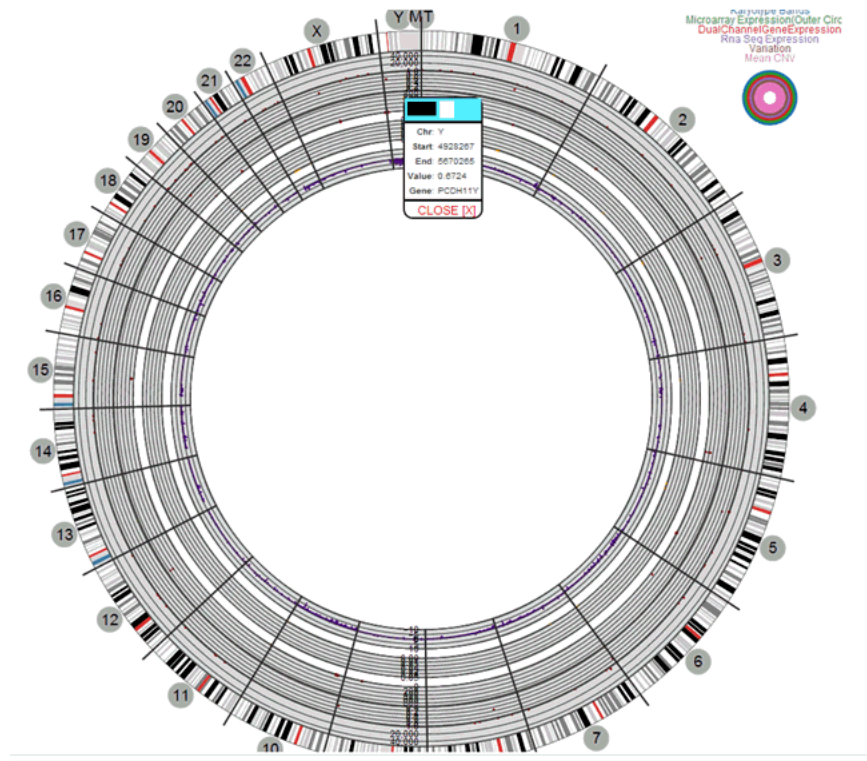
> **Note:**   For the circular viewer to run properly in IE9, turn off
> compatibility mode.

5.2.1 Selecting Data to Plot

5.2.1.1 Microarray Expression

The microarray expression panel displays all the specimens of the result type
expression. You can select only one specimen at a time.

For the selected specimen from the panel, you can view the list of hybridizations
available for that specimen. You can select maximum of two hybridizations from the
multiple choice box, for which the plot is rendered.

The color delimiter value is used to render color to the data points in the plot based on
the intensity value. If the intensity value is above the value defined in color delimiter
then the data points have green color Otherwise, the points have red color.

5.2.1.2 Sequencing-Variant Density

This option enables you to plot the density of variants for every 100-kb region for a
specimen. The sequence variant density panel displays all the specimens of the result
type sequence variants. You can select only one specimen at a time.

5.2.1.3 RNA-Seq Expression

This panel displays all the specimens of result type RNA Sequencing.You can select only one specimen at a time.

The color delimiter value is used to render color to the data points in the plot based on the RPKM value. If the RPKM value is above the value defined in color delimiter then the data points have green color. Otherwise, the points have red color.

You can use the check boxes provided next to each of the result types to determine if the specimen selected will be plotted or not. You can select one or all of the result types for plotting of the graph. Once the specimens are selected, the circular genomic plot is displayed. The outermost circle in the plot is the Cytoband. The supported Cytoband versions are: hg-18 and hg-19.

> **Note:**   To display the cytoband in the Circular Genomic plot we need an entry mapping each DNA Reference Version to the Cytoband in the TRC_LOOKUP_CODE table in the APP_schema.

For each reference version we must insert a row with the following values:

- CODE_TYPE : TRC_REFVERSION_CYTOBAND

- CODE:cytobandHG19 (for HG19), cytobandHG18 (for HG18), or cytobandHG38 (for HG38)

- CODE_NAME :Name of the loaded DNA reference version (for example, V69)

You can hover over the plot to get details such as:

- Microarray Expression: Chromosome, Start Position, End Position, Value and Gene

- Copy Number Variation: Chromosome, Start Position, End Position and Value

- Sequence Variants: Chromosome, start position, end position, and value. The density value is calculated by: Total number of variants for 100kb region / 100000

- Dual Channel Microarray Expression: Chromosome, Start Position, End Position, Value and Gene

- Rna Seq Expression: Chromosome, Start Position, End Position, Value and Gene

5.2.1.4 Copy Number Variation

The panel displays all the specimens of the result type copy number variation. You can select only one specimen at a time.

The value for copy number variation depends on the CNV result type selected. If the selected specimen contains data from Genome_Wide_SNP_6 array, then the value for CNV is taken from segment mean stored in the database. If the selected specimen has data from complete genomics, then the value is calculated based on the calledPloidy value stored in the database. The value to be plotted is calculated using the following formula for CNV data from complete genomics:

log2(called_ploidy/expected_ploidy)

where,

expected_ploidy is 2 for chr1-22

expected_ploidy is 2 for chrX for females.

expected_ploidy is 1 for chrX outside the pseudo-autosomal region in males

expected_ploidy is 2 for chrX inside pseudo-autosomal region in males

The pseudo autosomal regions on chrX for 'NCBI build 37' as reported by Complete Genomic are 60000 - 2,699,519 and 154,931,043 - 155,260,559.

The pseudo autosomal regions on chrX for 'NCBI build 36' as reported by Complete Genomic are 0 -2,709,519 and 154,584,237 - 154,913,753.

For called_ploidy zero, there the log2 will be infinity, in such cases the final value is taken as -2.

5.2.1.5 Dual Channel Microarray Expression

This panel displays all the specimens of result type Dual Channel Microarray expression. You can select only one specimen at a time.

The color delimiter value is used to render color to the data points in the plot based on the Log2Ratio value. If the Log2Ratio value is above the value defined in color delimiter then the data points have green color. Otherwise, the points have red color.

6

Genomic Query

This chapter contains the following topics:

6.1 Genomic Query Tab

The **Genomic Query** tab can be utilized as a search interface into the patient genomic results in Oracle Healthcare Omics. Two search modes are provided.

6.2 Using Gene Search

This tab lets you to specify one or more genes to be searched across results in OHO. To perform a gene search:

1. Navigate to **Genomic Query** > **Gene Search**.

2. To load a recently used query, click **Load Recent Query**. You can also search for the gene by gene name, pathway and gene set.

3. Enter the **Assembly Version**, **Species** and DNA Reference version.

   Only the annotations of the selected genes are retrieved based on the DNA reference version. For example, if you select Assembly Version as GRCh37, then the DNA Reference Versions associated with GRCh37 assembly is displayed. The default is the one that is set as Preferred Version in Oracle Healthcare Omics (OHO, formerly known as ODB). You can also select another DNA Reference Version.The DNA Reference Version facilitates retrieving information such as Gene Name in the report.

4. Click **Submit**.

Figure 6–1   Gene Search

The results are displayed on the right. It includes the following:

- Result Summary - this displays how many genes were searched for, how many have results found and how many do not have any results present in OHO for the criteria selected, what types of results are found if any.

- Result Details - Tabular list of metadata for each gene that is found to have genomic results in OHO. The columns include:

  - HUGO gene name,

  - Ensembl gene name,

  - Assembly Version (Reference Version),

  - the number of distinct patients (or subjects) found with any data for a given gene,

  - the number of distinct specimen found with any data for a given gene,

  - the number of matching result files present in OHO

  - the type of result files present in OHO. For example, VCF, MAF, gene expression.

5. To export the resulting table to a spreadsheet, click **Export**.

6. To print the results, click **Print**.

7. Click the Patient count to navigate to different screens like View Record, Circular Genomic Viewer, Cohort List, Cohort Timelines, Cohort Reports and Genomic Data export.

If the pathway, geneSet or genes have been selected from the ad hoc genes selection, the result details table will be paginated based on gene selection. The first page displays the result for 25 gene selection and the next page will be the result for next 25 genes and so on.

6.3 Using Variant Search

This tab lets you to specify one or more variant identifiers to be searched across results in Oracle Healthcare Omics (OHO, formerly known as ODB). To perform a variant search:

1. Navigate to **Genomic Query** > **Variant Search**.

2. To load a recently used query, click **Load Recent Query**. You can also search for the variant by variant ID or attributes.

   You can specify Genomic Position by typing in chr#:from-to location. For example, to find variants in base pair region 1-200 on chromosome 7, you would need to type in chr7:1-200. Also, if you want to search for any variants on chromosome 7, you can enter chr7.

   Additional options that can be specified include known or novel status on the variant, variant type such as insertion, deletion, substitutions, indel, complex. The strand is the directionality of the variant and it can be either plus (+) or minus (-).

3. Enter the **Species** and **Assembly Version**.

4. Click **Submit**.

   The results are displayed on the right. It includes the following:

   - Result Summary - this displays how many variants were searched for, how many have results found and, if applicable, how many do not have any results present in OHO for the criteria selected, what types of result files are found for the variants if any.

   - Result Detail - Tabular list of metadata for each variant that is found to have genomic results in OHO. The columns list includes the following option:

     – Variant Name, if the name exists,

     – Location of the variant,

     – Hugo Name

     – Assembly Version (Reference Version)

     – Status, whether the variant is known or novel,

     – the replace tag that tells the end user what exactly the variant is,

     – Tag, whether this variant is an insertion, deletion and so on,

     – Impact, the impact of a given variant (only available for known variants),

     – Study, the study this variant has been found to have results for,

     – Number of Patients, the number of distinct patients found with any data for the particular variant,

     – Number of Specimen, the number of distinct specimen found with any data for the particular variant

     – Number of Result File, the count of type of result files present in OHO,

     – Result File Type, type of result files present in OHO, for example, VCF, MAF, gene expression with data for a given variant.

> **Note:** On gene search selections: A variant may be present in alternate genes that overlap with the gene region of the selected gene in a given version. For any variant found, these alternate gene annotations are displayed to the user along with annotations from gene of interest.

5. To export the data to a spreadsheet, click **Export**.

6. To print the current screen, click **Print**.

7

OBIEE Reports

This chapter describes the OHTR reports in the analytics application. It contains the following topics:

- Section 7.1, "Viewing Standard Reports"

- Section 7.2, "Selection of Date Range Using Statistics from and to Prompt"

- Section 7.3, "Patients Diagnosed and Treated"

- Section 7.4, "Number of Treatments Performed"

- Section 7.5, "Ethnicity or Race of Newly Diagnosed Patients"

- Section 7.6, "Age or Gender of Newly Diagnosed Patients"

- Section 7.7, "Biospecimen Samples Status"

- Section 7.8, "Details of Each Report and Prompt Set"

> **Note:** To enable context-sensitive help in OBIEE reports, see Deploying Help Files for OBIEE Dashboard in the *Oracle® Healthcare Translational Research Installation Guide*.

7.1 Viewing Standard Reports

OBIEE has seven preconfigured OHTR reports. To view these reports:

1. Log into the OBIEE Analytics application and open the Cohort Explorer Dashboard.

2. Enter a date range in the **Statistics from** and **to** fields.

3. Click **Apply**. The reports for the selected date range are displayed.
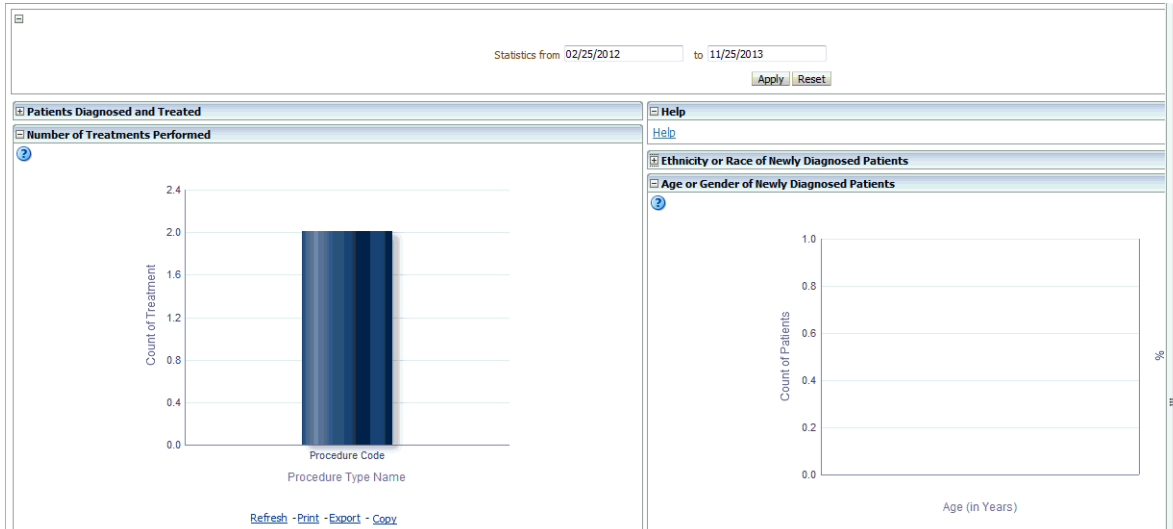
7.2 Selection of Date Range Using Statistics from and to Prompt

By default, all reports display the past one month's data from the current date. You can also view data for a date range of your choice.

For more information, see the prompt details in Statistics from-to on page 4.

7.3 Patients Diagnosed and Treated

The Patients Diagnosed and Treated report displays the past patient data for the user-defined date range, which by default displays one month's data from the current date.

It is preconfigured during the initial setup or installation phase and can be altered only by a CE-Administrator or CE-Developer. The columns of the resulting patient count table are:

- Number of diagnosed patients: Number of patients diagnosed with a specific disease during the time period specified.

- Number of patients successfully treated: Number of patients that have been treated successfully during this time period.

- Number of consented patients: Number of patients that have consented to specific treatments (type of consent considered is customized based on your requirements) during the specified time period.

- Number of patients undergoing treatment: Number of patients that are or have been undergoing any type of treatment during the specified period of time.

For more information, see the report details in Patients Diagnosed and Treated

7.4 Number of Treatments Performed

The Number of Treatments Performed report counts the number of treatments performed during the specified from-to time period. The default being the past one month's data from the current date.

Treatments are categorized according to types and include Surgery, Chemotherapy, Radiation, and Gene Therapy and so on. Treatments are selected during the configuration phase by either the CE-Administrator or CE-Developer and cannot be modified by other users.

For more information, see the report details in Number of Treatments Performed

7.5 Ethnicity or Race of Newly Diagnosed Patients

The Ethnicity or Race of Newly Diagnosed Patients report is a pie chart view of how many patients of a particular race or ethnicity have been diagnosed with a disease during the specified time period. The default being the past one month's data from the current date. You can toggle between Ethnicity (by default) and Race to see the counts and percentages of patients.

For more information, see the report details in Ethnicity or Race of Newly Diagnosed Patients

7.6 Age or Gender of Newly Diagnosed Patients

The Age or Gender of Newly Diagnosed Patients report displays how many patients within a particular age category and gender have been diagnosed in the facility during the user specified time period. The default being the past one month's data from the current date. There following five age categories are configured:

- 0-18 years old
- 19-35
- 36-50
- 51-65
- 66+

For more information, see the report details in Age or Gender of Newly Diagnosed Patients

7.7 Biospecimen Samples Status

The Biospecimen Samples Status report displays the counts of specimen samples that are collected during the specified time period. The default being the past one month's data from the current date. The counts are broken down by the Anatomical Site, listed along the rows.

The list of available sites is configured by the customer during the system installation. The Anatomical Site is a hierarchy where you can drill into the details of each Anatomical site to view its subcategories along with associated counts. The display columns are:

- Type of specimen: Multiple types of specimen collected for each anatomical site. For example, tissue, cell culture, liquid, and so on. All of the available types of specimen are listed for each site.

- Number of Specimen Collected in Facility: Displays the number of specimens collected for each listed Anatomical Site and Type during the specified time period. The default being past one month's data from the current date.

- Number of Patients Corresponding to Collected Specimen: Number of unique patients that correspond to the specimen collected in the previous column.

- Number of Patients Providing Specimen who Consented: Number of patients that have signed consent forms allowing usage of their samples for research. The exact type of consent considered vary by the facility and is configured during the initial system installation.

For more information, see the report details in Biospecimen Samples Status

7.8 Details of Each Report and Prompt Set

7.8.1 Statistics from-to

You are prompted for a date range, where the default value is the past one month's data from the current date. This prompt is used to filter all reports on the dashboard.

Audience
- Physician

- Researcher

- Clinician

- Limited Access User

Prompt Set Type
- Text Field

Location
- CE Dashboard

Dimensions
- Diagnosis.Diagnosis Start Date

- Diagnosis End Date, Procedure. Procedure Start Date

- Procedure. Procedure End Date

- Specimen.Specimen Collection Date

- Medication. Medication Start Date

- Medication.Medication End Date

- Consent.Consent Start Date, Consent. Consent End Date

- Specimen.Anatomical Site Name

Supplementary Reports
None

Reports Referenced

- TRC-CER-001

- TRC-CER-003

- TRC-CER-004

- TRC-CER-005

- TRC-CER-006

Reports Referencing This Prompt

None

Prompt Descriptions

Table 7–1 describes the prompts in the Statistics from-to Prompt Set.

Table 7–1   Statistics from-to (Prompt Set)

Prompt Page Number, Prompt Page Heading	Prompt Heading	Prompt Type or Default Value Layer	Measure or Dimension in PL Layer
1, NA	Statistics from to	Text Field or FROM 1 month back from current date TO current date	Diagnosis.Diagnosis Start Date, Diagnosis End Date, Procedure. Procedure Start Date, Procedure. Procedure End Date, Specimen.Specimen Collection Date, Medication. Medication Start Date, Medication.Medication End Date, Consent.Consent Start Date, Consent. Consent End Date

7.8.2 Patients Diagnosed and Treated

This report displays statistics on patients in a facility for the time period specified, the default being the past one month's data from the current date. These statistics include the number of Patients Diagnosed and Treated with disease during the time interval and how many have given consent to medical care.

Audience

- Physician

- Researcher

- Clinician

- Limited Access User

Report Type

- Table

Location

- CE Dashboard

Dimensions

- Diagnosis

- Treatment

- Consent

Supplementary Prompts

- Statistics from-to

Reports Referenced

None

Reports Referencing this Report

None

Column Descriptions

Table 7–2 describes the columns in the Patients Diagnosed and Treated report.

Table 7–2   Patients Diagnosed and Treated)

Table Heading	Column Heading	Measure or Dimension in Presentation Layer
NA	Number of Diagnosed Patients	TRC-CEM-003
NA	Number of Patients Successfully Treated	TRC-CEM-006
NA	Number of Patients who Consented	TRC-CEM-005
NA	Number of Patients Undergoing Treatment	TRC-CEM-007

7.8.3 Number of Treatments Performed

This report displays statistics on the number and types of procedures or treatments that have been performed during specified time period. The default being the past one month's data from the current date.

Audience

- Physician

- Researcher

- Clinician

- Limited Access User

Report Type

- Graph: Vertical Bars

Location

- CE Dashboard

Dimensions

- Treatment

Supplementary Prompts

- Statistics from-to

Reports Referenced

None

Reports Referencing This Report

None

Column Descriptions

Table 7–3 describes the columns in the Number of Treatments Performed report.

Table 7–3   Patients Diagnosed and Treated

Table Heading	Column Heading	Measure or Dimension in Presentation Layer
NA	Surgery	TRC-CEM-010
NA	Chemotherapy	TRC-CEM-011
NA	Radiation	TRC-CEM-012
NA	Targeted (Gene Therapy)	TRC-CEM-013

7.8.4 Ethnicity or Race of Newly Diagnosed Patients

This report displays the ethnicity and race statistics for the patients in the facility who have been diagnosed during the time period specified in the dashboard prompt. The default being the past one month's data from the current date. The total count of patients in this report is based on the patients diagnosed during the period specified by the user.

Audience

- Physician

- Researcher

- Clinician

- Limited Access User

Report Type

- Graph: Pie Type: Default Style: Default

Location

- CE Dashboard

Dimensions

- Patient.Ethnicity

- Name Patient. Race Name

Supplementary Prompts

- Statistics from-to Display: Ethnicity, Race

Reports Referenced

None

Reports Referencing This Report

None

Column Descriptions

Table 7–4 describes the columns in the Ethnicity or Race of Newly Diagnosed Patients report

Table 7–4   Ethnicity or Race of Newly Diagnosed Patients

Table Heading	Column Heading	Measure or Dimension in Presentation Layer
NA	% Total Patients	TRC-CEM-015
NA	Number of Diagnosed Patients	TRC-CEM-014

7.8.5 Age or Gender of Newly Diagnosed Patients

This report displays the age and gender statistics for patients in the facility who have been diagnosed during time period specified in the dashboard prompt. The default being the past one month's data from the current date. The total count of patients in this report is based on the patients diagnosed during the period specified by you.

Audience

- Physician

- Researcher

- Clinician

- Limited Access User

Report Type

- Graph: Pie Type: Default Style: Default

Location

- CE Dashboard

Dimensions

- Patient.Ethnicity

- Name Patient. Race Name

Supplementary Prompts

- Name Patient. Race Name

Reports Referenced

None

Reports Referencing This Report

None

Column Descriptions

Table 7–5 describes the columns in the Age or Gender of Newly Diagnosed Patients report

Table 7–5   Age or Gender of Newly Diagnosed Patients

Table Heading	Column Heading	Measure or Dimension in Presentation Layer
NA	% Total Patients	TRC-CEM-017
NA	Number of Diagnosed Patients	TRC-CEM-016
NA	Male	Patient.Gender
NA	Female	Patient.Gender
0-18	NA	Patient.Age
19-35	NA	Patient.Age
36-50	NA	Patient.Age
51-65	NA	Patient.Age
66+	NA	Patient.Age

7.8.6 Biospecimen Samples Status

This report displays the statistics on biospecimen collected during the time period specified. The default being the past one month's data from the current date. These statistics include the number of samples collected, the types of samples and their anatomical site, the number of patients corresponding to these specimens and the number of patients that consented for these specimens to be utilized.

Audience

- Physician
- Researcher
- Clinician
- Limited Access User

Report Type

- Table

Location

- CE Dashboard

Dimensions

- Specimen.Specimen Type Name
- Specimen.Anatomical Site Name

Supplementary Prompts

- Statistics from-to (Prompt Set TRC-CEP-001)

Reports Referenced

None

Reports Referencing this Report

None

Column Descriptions

Table 7–6 describes the columns in the Biospecimen Samples Status report

Table 7–6   Biospecimen Samples Status

Table Heading	Column Heading	Measure or Dimension in Presentation Layer
NA	Anatomical Site	Specimen.Anatomical Site Name
NA	Type of Specimen	Specimen.Specimen Type Name
NA	Number of Specimen Collected in Facility	TRC-CEM-018
NA	Number of Patients Corresponding to Collected Specimen	TRC-CEM-019
NA	Number of Patients providing Specimen who Consented	TRC-CEM-020

8

User Management

This section contains the following topics:

- User Accounts
- Roles and Permissions
- Working with User Groups

8.1 User Accounts

In the current release of OHTR, the user has permissions to make a few changes to basic accessibility options. However, the viewing screens cannot be customized.

If Oracle Access Manager (OAM) is configured, the user has the following functions:

1. The user can log into the OHTR application by using the single sign-on interface, which can be shared among multiple applications.

   For example, with an Oracle Business Intelligence Enterprise Edition Plus (OBIEE) full license, the same credentials can be used for generating OBIEE reports.

2. The user creates the password, which is not visible to an administrator.

3. After a configurable number of unsuccessful login attempts, the user is locked out.

4. After a configurable amount of inactive time, the login session times out.

5. Roles are automatically setup as described in Section 8.2, "Roles and Permissions".

If OAM is not configured, the identified roles must be manually set up in a WebLogic instance. For more information, see Create Users and Add Users to Groups in the *Oracle® Fusion Middleware Oracle WebLogic Server Administration Console Online Help* available at

https://docs.oracle.com/middleware/1221/wls/WLACH/taskhelp/security/Manage UsersAndGroups.html

You may need to perform policy migration in OBIEE to see the roles. For details on policy migration, see the *Oracle® Healthcare Translational Research Installation Guide*.

8.2 Roles and Permissions

The OHTR user is assigned one or more of the following roles:

Role	Permissions	Accessible Screens
trc-bioinformatician-group	■ Download files that have a link stored in Oracle Healthcare Omics (OHO, formerly known as ODB) and can be located in the middle tier in an accessible location. ■ Cannot build new reports or modify the content of existing reports.	■ Cohort Query Tab ■ Cohort Query Tab: Genomic Data (tab in accordion) ■ Query Tab: Relative Time Events: gene variant (radio button) ■ Cohort Viewer (top tab) ■ Cohort Viewer: Cohort List ■ Cohort Viewer: Cohort Timeline ■ Cohort Reports ■ Cohort Viewer: Genomic Data Export ■ Single Patient Viewer: View Record ■ Single Patient Viewer: View Record: Genomic Data Collected ■ Circular Genomic Viewer (Visquick) ■ Genomic Query ■ Genomic Query: Columns after searching for gene/variant coming from CDM (Patient Count, Specimen Anatomical Site columns) ■ My Workspace: My Recent Queries ■ My Cohort Lists ■ Queries or Lists shared with me ■ Queries or Lists shared with All ■ My Queries ■ Gene Sets ■ Manage Gene Sets ■ Jobs
trc-cohort-group	■ Query any data from CDM but cannot query OHO directly. ■ Export the clinical data in a supported format and view the Dashboard. ■ Cannot build new reports or modify the content of existing reports.	■ Cohort Query Tab ■ Cohort Query Tab: Relative Time Events: gene variant (radio button) ■ Cohort Viewer (top tab) ■ Cohort Viewer: Cohort List ■ Cohort Viewer: Cohort Timeline ■ Single Patient Viewer: View Record ■ My Workspace: My Recent Queries ■ Short Cuts ■ My Cohort Lists ■ Queries or Lists shared with me ■ Queries or Lists shared with All ■ My Queries
trc-limited-user-group	■ View the patient count on the Query Patients page ■ Cannot build new reports or modify the content of existing reports.	■ Cohort Query Tab (Save Query button is not accessible) ■ My Workspace: My Recent Queries

Role	Permissions	Accessible Screens
trc-omics-group	- Query and read data from OHO - Export omics data in file formats readable by genome viewers. For example, VCF, SEG, RES formats for IGV. - Cannot build new reports or modify the content of existing reports.	- Genomic Query - My Workspace: My Recent Queries - Short Cuts - Gene Sets - Jobs
trc-comics-limited-user-group	- Query data from CDM and OHO using Query Patients interface - Can only access Patient Genomic Data export within the Patient Viewer interface - Create Gene Sets - View queries and gene sets	- Cohort Query Tab - Cohort Query Tab: Genomic Data (tab in accordion) - Query Tab: Relative Time Events: gene variant (radio button) - Cohort Viewer (top tab) - Cohort Viewer: Genomic Data Export - My Workspace: My Recent Queries - Gene Sets - Jobs

Role	Permissions	Accessible Screens
trc-pi-user-group	■ Specific privileges allowing access to identifiable information on patients or subjects ■ In Subject context, can view merged version of Patient and Subject clinical data all in one View Record page.	■ View personally identifiable (PI) information on the Cohort List page. ■ View PI data in Single Patient or Subject Viewer > View Record page. ■ Cohort Query Tab ■ Cohort Query Tab: Genomic Data (tab in accordion) ■ Query Tab: Relative Time Events: gene variant (radio button) ■ Cohort Viewer (top tab) ■ Cohort Viewer: Cohort List ■ Cohort List (PI attributes) ■ Cohort Viewer: Cohort Timeline ■ Cohort Reports ■ Cohort Viewer: Genomic Data Export ■ Single Patient Viewer: View Record ■ Single Patient Viewer: View Record (PI attributes) ■ Single Subject Viewer: View Record (PI attributes) ■ Single Subject Viewer: View Record (show patient data together) ■ Single Patient Viewer: View Record: Genomic Data Collected ■ Circular Genomic Viewer (Visquick) ■ Genomic Query ■ Genomic Query: Columns after searching for gene/variant coming from CDM (Patient Count, Specimen Anatomical Site columns) ■ My Workspace: My Recent Queries ■ Short Cuts ■ My Cohort Lists ■ Queries or Lists shared with me ■ Queries or Lists shared with All ■ My Queries ■ Gene Sets ■ Saved Queries ■ Jobs
trc-admin-group	■ Create new user groups and so on for sharing cohort queries or lists ■ Add or remove users from groups ■ Manage user roles	Manage User Group

Users belonging only to the following groups have limited functionality based on the assigned roles:

■ trc-comics-limited-user-group

- trc-limited-user-group
- trc-basic-user-group

8.3 Working with User Groups

To simplify sharing, create user groups and assign users to them. Instead of sharing queries or lists with each user individually, set up a list of users to share multiple items.

Note that only the user that is the owner of a given list can modify members of a given list. However, any user can elect to utilize a given list for sharing and see members of a list, even if he or she is not the owner.

8.3.1 Creating a User Group
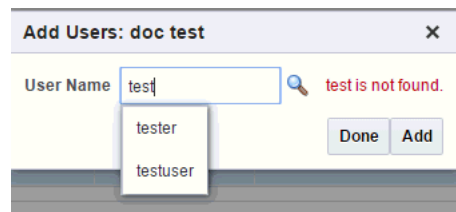
To create a user group:

1. Navigate to **Manage User Groups** > **User groups**.

2. Click **Create**. The Create User Group box is displayed.

3. Enter a name and description for the user group.

4. Click **Submit**.

8.3.2 Assigning Users to a Group

To assign users to a group:

Immediately After you Create a Group

1. Navigate to **Manage User Groups** > **Group has members**.

2. Click **Add**. The Add Users box is displayed.

3. Enter the user name you want to add. A drop-down list displays the matching choices.

4. Select the names you want to add.

5. Click **Add**. The selected users are added to the group.

6. Click **Done**.

Searching for a Group you Created Earlier

1. Navigate to **Manage User Groups** > **Search User Groups**.

2. Enter the details of the group you want to find.

3. Click **Search**. The user group and its details are displayed.

4. Navigate to **Manage User Groups** > **Group has members**.

5. Click **Add**. The Add Users box is displayed.

6. Enter the user name you want to add. A drop-down list displays the matching choices.

7. Select the names you want to add.

8. Click **Add**. The selected users are added to the group.
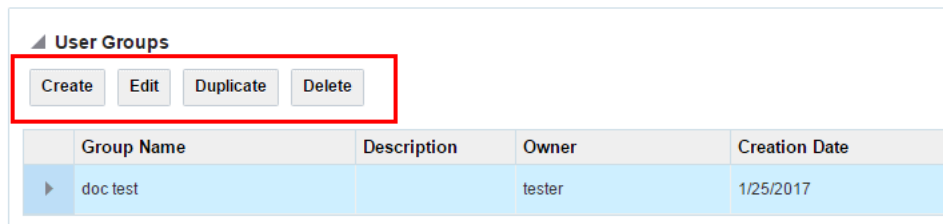
9. Click **Done**.

8.3.3 Removing Users from a Group

To remove users from a group:

1. Navigate to **Manage User Groups** > **Search User Groups**.

2. Enter the details of the group you want to find.

3. Click **Search**. The user group and its details are displayed.

4. Navigate to **Manage User Groups** > **Group has members**.

5. Select the user you want to remove.

6. Click **Remove**.

8.3.4 Managing User Groups

You can edit, duplicate or delete a user group using the buttons at the top of the **User Groups** section.

8.3.4.1 Editing a User Group

To edit the name and description of a user group:

1. Navigate to **Manage User Groups** > **Search User Groups**.

2. Enter the details of the group you want to find.

3. Click **Search**. The user group and its details are displayed in the User Groups section.

4. Select the user group you want to modify.

5. Click **Edit**.

6. The **Edit User Group** box is displayed. Modify the name and description of the gene set.

7. Click **Submit**.

8.3.4.2 Duplicating a User Group

To edit the name and description of a user group:

1. Navigate to **Manage User Groups** > **Search User Groups**.

2. Enter the details of the group you want to find.

3. Click **Search**. The user group and its details are displayed in the User Groups section.

4. Select the user group you want to duplicate.

5. Click **Duplicate**.

6. The **Duplicate User Group** box is displayed. Modify the name and description of the user group.

7. Click **Submit**.

8.3.4.3 Deleting a User Group

To delete a user group:

1. Navigate to **Manage User Groups** > **Search User Groups**.

2. Enter the details of the group you want to find.

3. Click **Search**. The user group and its details are displayed in the User Groups section.

4. Select the user group you want to delete.

5. Click **Delete**. A confirmation box is displayed.

6. Click **Submit**. The user group is deleted.